

El estudio observacional

Mirta Ciocca¹  · Ricardo Mastai¹  · Arturo Cagide² 

¹ Hospital Alemán de Buenos Aires.

² Hospital Italiano de Buenos Aires.

Ciudad Autónoma de Buenos Aires, Argentina.

Acta Gastroenterol Latinoam 2025;55(2):70-78

Recibido: 04/05/2025 / Aceptado: 02/06/2025 / Publicado online: 30/06/2025 / <https://doi.org/10.52787/agl.v55i2.489>

Una historia de ficción

En un centro especializado en enfermedades hepáticas se ha estado empleado un nuevo fármaco (“*Tto. X*”) para el tratamiento de la hipertensión portal. Los profesionales están convencidos de que la intervención reduce la hemorragia digestiva asociada a este síndrome clínico.

Realizado un análisis estadístico demuestran que con el *Tto. X*, efectivamente hubo una disminución significativa de aquella complicación, con una $p < 0,05$.

Pero surge un interrogante: ¿habrán recibido el tratamiento solamente los pacientes con mejor condición clínica y pronóstico favorable, como por ejemplo menor edad, menor grado de hipertensión portal, o ausencia de plaquetopenia? Si así fuera, ¿la $p < 0,05$, se debe al tratamiento o a que la intervención actuó como una “seleccionadora” de una población de menor riesgo?

El diseño observacional

La Figura 1 ilustra un estudio observacional ficticio, donde la variable de evolución, **punto final del estudio o “outcome”**, podría estar asociada a la variable de exposición (objetivo de la investigación) pero también a un cierto número de **confundidores** (o covariables) que pueden condicionar aquella asociación.

Por ejemplo, como se adelantó, si una intervención puede modificar favorablemente el **outcome** -para el caso mortalidad de causa hepática o hemorragia digestiva no mortal- se indica preferentemente a pacientes con mayor número de plaquetas o una enfermedad de menor severidad, ese beneficio podría deberse al tratamiento, pero también a que el procedimiento “seleccionó” individuos con menor riesgo de hemorragia. Obsérvese que la condición detallada requiere la asociación de los confundidores con la intervención.

En otras palabras, todo **desbalance** en la prevalencia de confundidores entre los grupos en comparación puede

Correspondencia: Mirta Ciocca

Correo electrónico: cioccamirta@gmail.com

alterar la asociación objetivo del estudio. De modo que, cuando se considera solo la relación entre intervención y *outcome* mediante un análisis **univariado** o **no ajustado**, la *p* resultante podría no tener valor clínico en cuanto al efecto del tratamiento.

En principio, la situación descripta no afecta a los ensayos randomizados (experimentales) ya que la indicación de intervenir no resulta de una decisión médica sino del azar, de modo que los confundidores se distribuyen en forma balanceada en ambos grupos, con y sin intervención, por lo que la asociación exposición / *outcome* se "independiza" de sus efectos. Sin embargo, el azar puede fallar en ocasiones, especialmente si la muestra es de tamaño reducido.

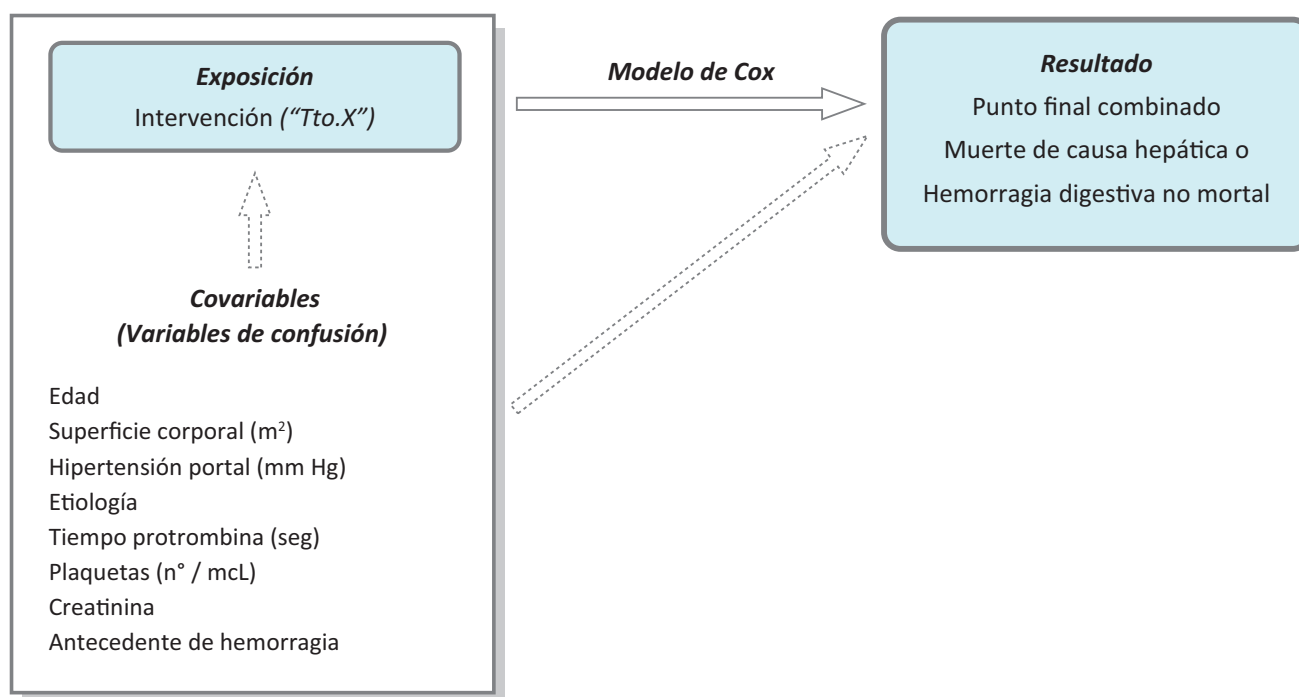
El procedimiento habitual para resolver el problema en un ensayo de seguimiento es el método **multivariado**

de regresión logística o tiempo al evento de riesgos proporcionales de Cox (de aquí en más "modelo de Cox").

Con esta metodología se analizan en conjunto, la variable de exposición y los confundidores (seleccionados por su asociación estadística aislada o el aporte bibliográfico (**variables independientes**)). (Figura 1). Este procedimiento es habitualmente referido como **ajuste**. Si en el análisis multivariado la exposición se asocia con significación estadística al *outcome* (**variable dependiente**) se concluye que dicha asociación no está condicionada a confundidores.

El análisis multivariado requiere una cierta proporción entre el número de variables independientes y la prevalencia/incidencia de la variable de evolución o *outcome*, lo que en ciertas condiciones limita las posibilidades para aplicar esta metodología.

Figura 1. Estudio multivariado



Se grafica la variable de exposición (intervención), y su posible asociación con la variable de evolución o *outcome* (áreas sombreadas, flecha de trazo continuo). Si las variables de confusión están asociadas tanto a la exposición como al *outcome* (flechas de trazo punteado) se podría atenuar o diluir la asociación, exposición / *outcome*. El análisis multivariado del conjunto (cuadro sombreado) estima la asociación "ajustada" de la exposición con el *outcome* independiente de los confundidores.

Score de Propensión ("Propensity Score", PS)

Una segunda metodología estadística es igualar los confundidores de modo que resulten balanceados en los grupos con intervención y control.

La Figura 2 ilustra el procedimiento. Como primer paso se estima estadísticamente la asociación entre los

confundidores (ahora variables independientes) y la exposición (ahora variable dependiente), empleándose un análisis multivariado (para el caso regresión logística).

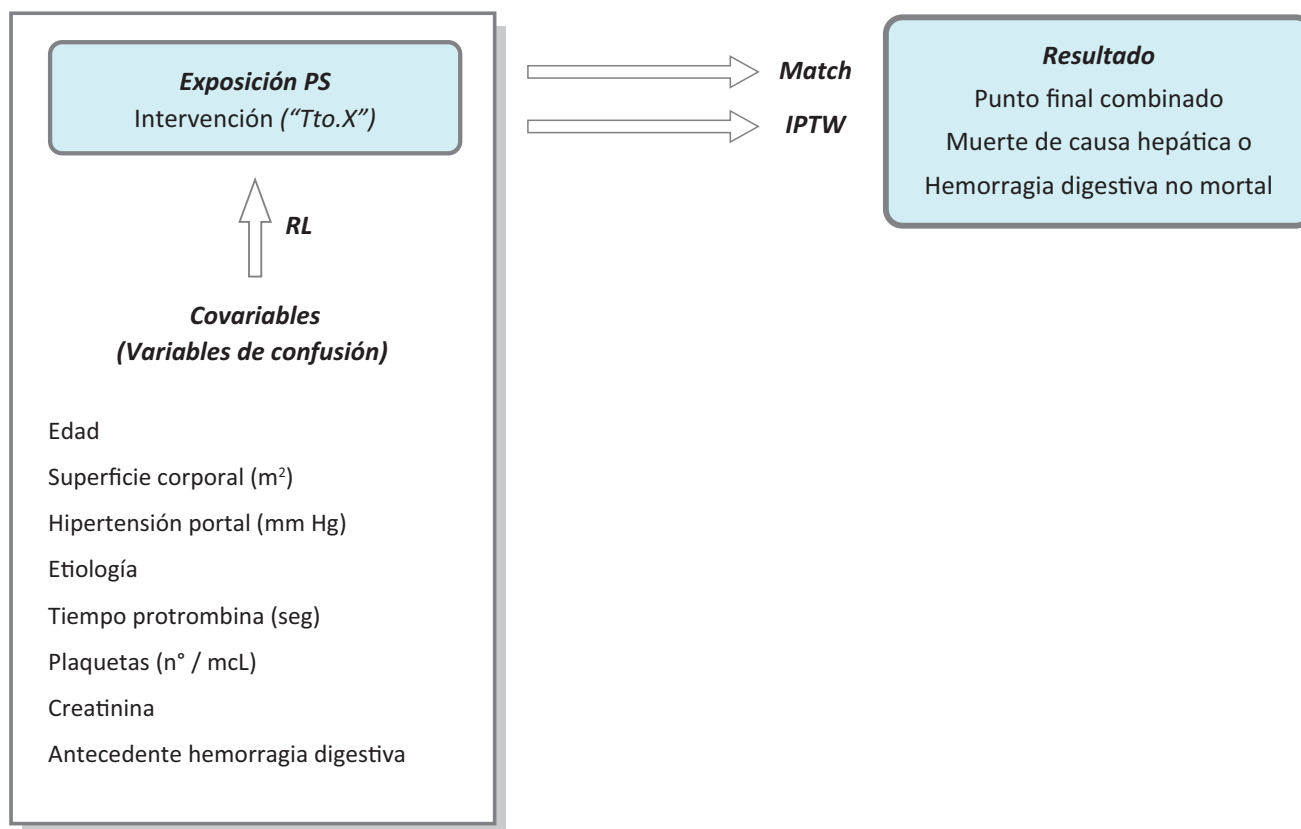
De su resultado se deriva el **Propensity Score** (PS) que es la probabilidad de quedar expuesto a la intervención independientemente de los confundidores. Los indivi-

duos con similar PS deberían tener igual probabilidad de recibir el tratamiento en investigación, hayan sido tratados o no con la intervención motivo del estudio.

Una ventaja que ofrece el PS sobre el análisis multi-

variado es que el número de las variables independientes no será limitado por la prevalencia del *outcome*, ya que la intervención, a diferencia del *outcome*, siempre tendrá suficiente número de observaciones.

Figura 2. Propensity score y análisis derivados



El objetivo final del estudio es, como en la Figura 1, estimar la posible asociación de la exposición, (intervención, "Tto.X") con el punto final (*outcome*). Previamente se analiza la asociación de los confundidores con la exposición y se calcula el *propensity score* (PS). Desde allí, el PS permite, aplicando diferentes metodologías estadísticas, como el emparejamiento ("match") o el inverso de la probabilidad ponderada de tratamiento (*IPTW*) estimar la asociación independiente de la exposición con el *outcome*. RL: Regresión Logística.

Variables de exposición y variables de evolución (*outcome*)

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

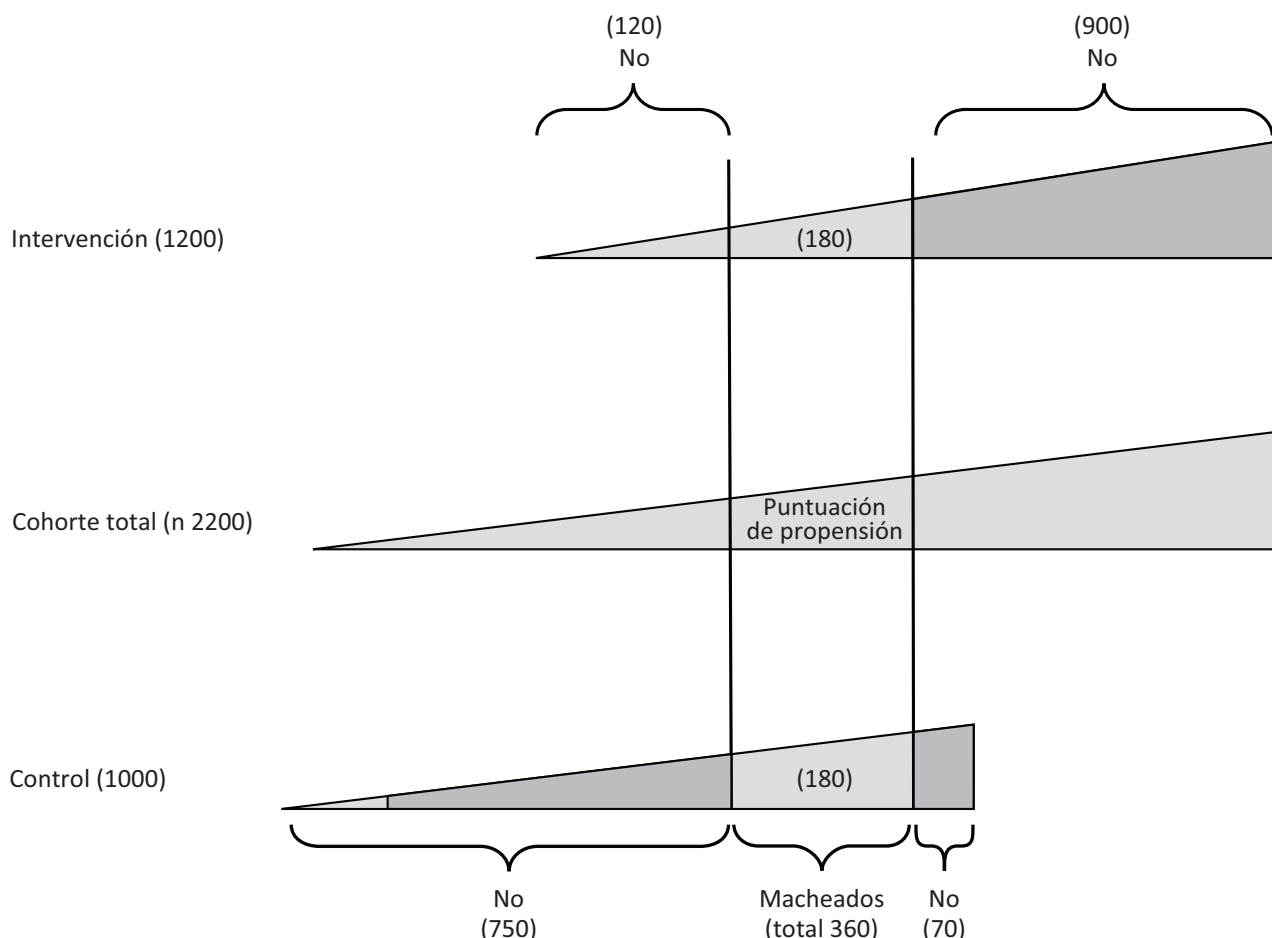
Como en cualquier análisis multivariado, el problema es que solo se consideran las variables independientes conocidas y disponibles y como resultado de esa situación el PS puede presentar defectos en la predicción de la exposición al tratamiento o intervención.

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

PS, ajuste de covariables y estimación del *outcome*

Apareamiento

Siguiendo la metodología anterior cada individuo quedará caracterizado, según sus características basales o confundidores, por un determinado PS. Algunos individuos del grupo intervención tendrán similar PS que los del grupo control, de modo que es posible aparear pacientes de ambos grupos según su PS. (Figura 3).

Figura 3. Apareamiento

Ejemplo teórico de un estudio de 2200 pacientes, 1200 con la intervención y 1000 controles. Se grafica, en una escala de valores crecientes, el PS de toda la muestra (centro), de los individuos intervenidos (superior) y de los controles (inferior). Cada individuo con intervención se aparea con un control con igual o muy cercano PS. Se integran 180 pares; la muestra se reduce a 16% del número original. (360/2200). En la muestra apareada los confundidores de ambos grupos, intervención y control, quedan "balanceados".

Ahora bien, un número significativo de individuos de ambos grupos, tratados y no tratados, serán excluidos al no disponerse del correspondiente par. Ese número de excluidos se relaciona en forma directa con el grado de desbalance de confundidores entre ambos grupos de la muestra original del estudio. De este modo, la conclusión del ensayo y su traslado a la práctica, se limita exclusivamente a la muestra emparejada, sin poder generalizarse a la totalidad de la población.

Ponderación inversa de la probabilidad de tratamiento (IPTW)

Este procedimiento, a diferencia del apareamiento, incluye la totalidad de la muestra en estudio.

Ahora bien, mientras que con el apareamiento el ajuste

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

La Figura 4 es un ejemplo teórico que compara el grupo de intervención con el grupo control. El número de plaquetas, dicotomizada en $>10^5$ mcL y $\leq 10^5$ mcL, es el único confundidor considerado para este ejemplo.

Hay un desbalance evidente ya que los individuos en el grupo tratado con plaquetas $>10^5$ son tres y solo uno en el grupo control. Se debe ajustar la covariable plaquetas para que ambos grupos puedan ser comparados en cuanto a un determinado *outcome*, por ejemplo, mortalidad de causa hepática o hemorragia digestiva no mortal. Para ello se estima el PS en cada uno.

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

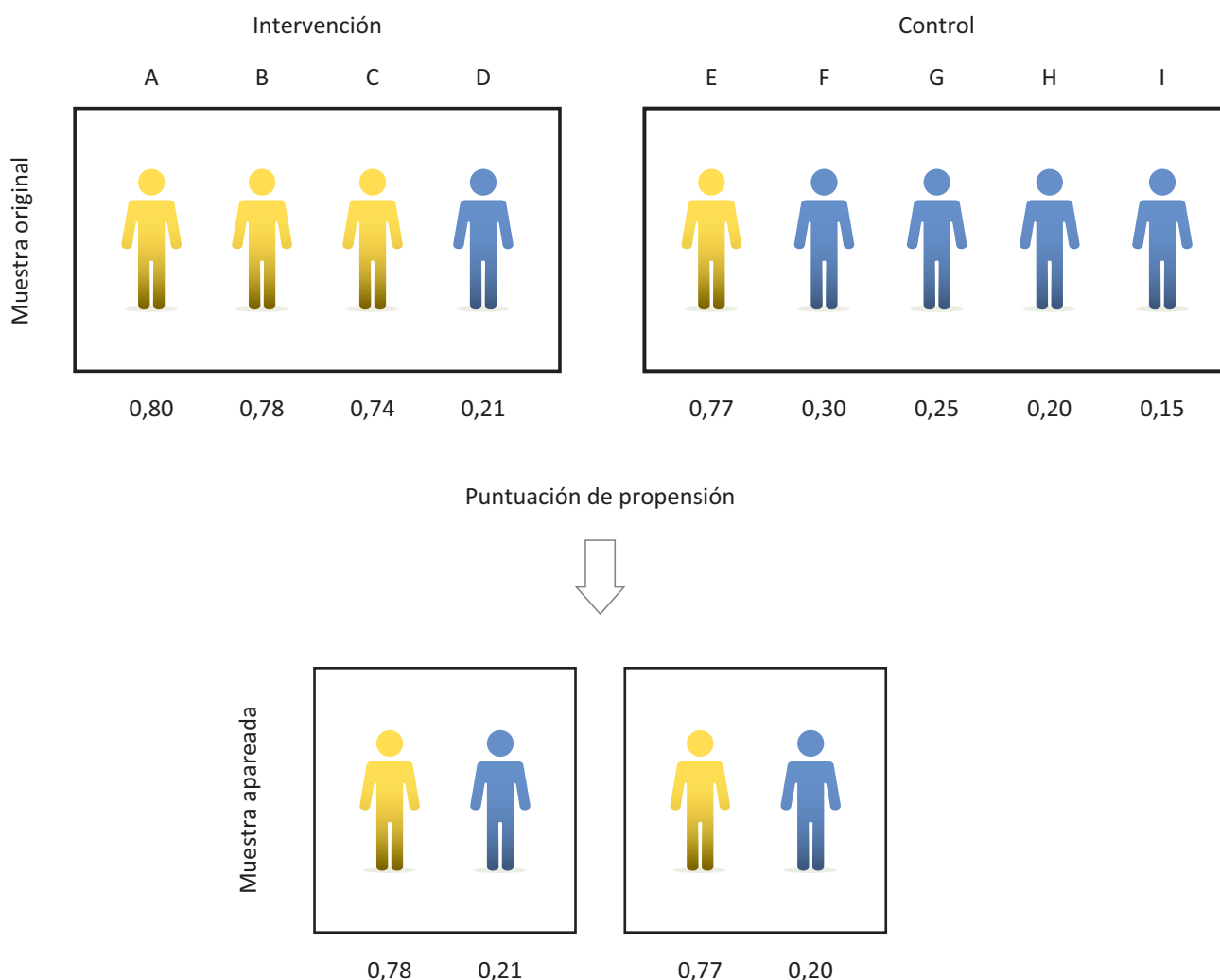
Si en el mismo ejemplo se emplea el IPWT el ajuste de la covariable plaquetas se alcanza, como ya se adelantó, incrementando artificialmente el número de observaciones mediante el procedimiento detallado en la Figura 5.

En el cálculo se emplea también el PS, pero en este caso, al no reducir el número de observaciones, se puede generalizar el resultado de la investigación a una población más representativa de la práctica clínica habitual.

Ahora bien, en un contexto real, la probabilidad de tratamiento queda condicionada a múltiples confundidores cuyo efecto global está representado en el PS, el cual será diferente para cada individuo. (Figura 6). El procedimiento, similar al explicado a propósito de la Figura 5, si bien reduce el desbalance, no es perfecto, persistiendo ciertas diferencias en la distribución de los individuos con plaquetopenia entre los grupos con y sin tratamiento, lo cual se debe al efecto de los otros confundidores.

La estrategia detallada a propósito de las plaquetas debe aplicarse a todas las variables o confundidores disponibles contemplados al calcular el PS.

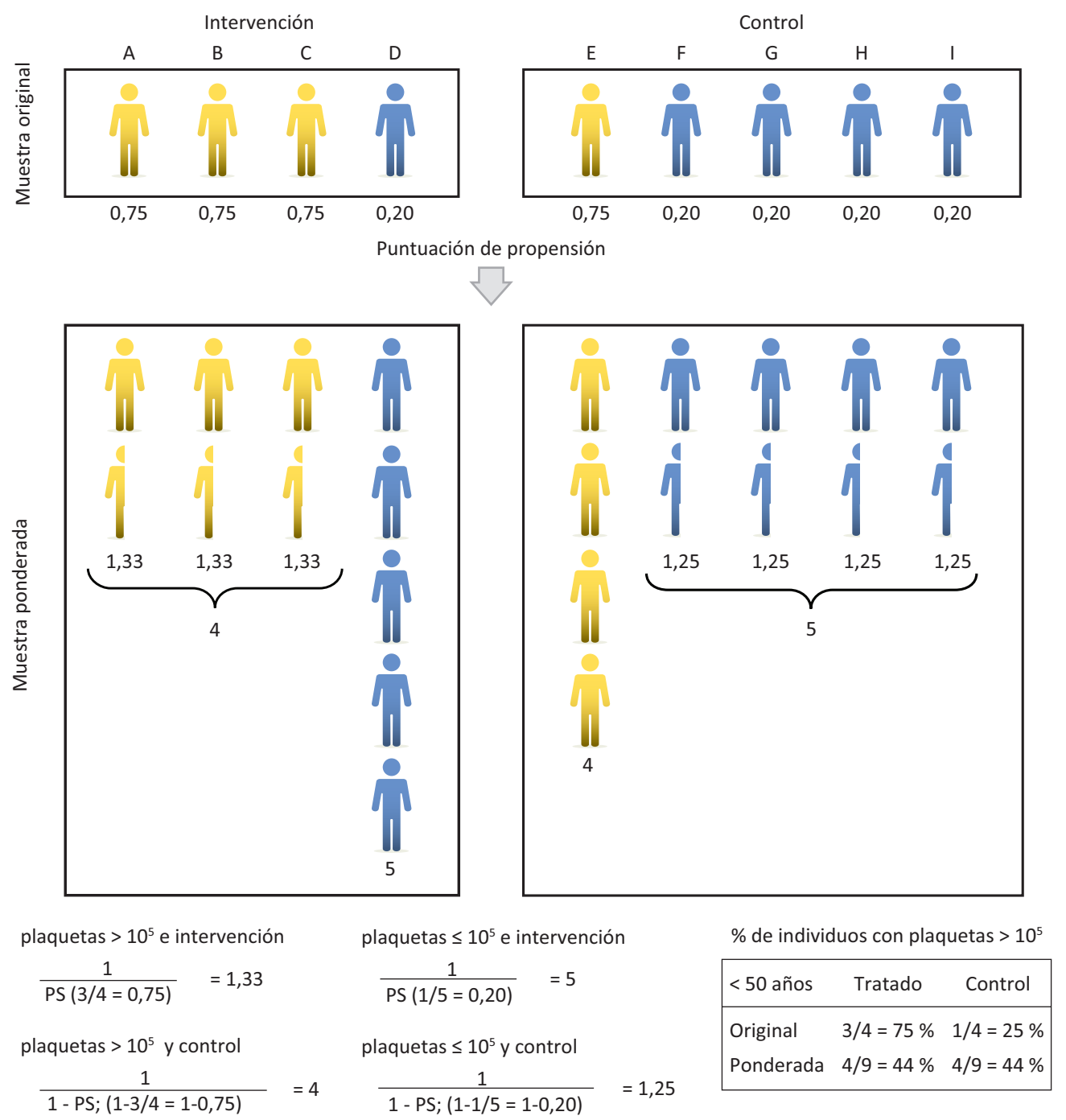
Figura 4. Apareamiento



Ejemplo teórico de un estudio observacional que compara la intervención (4 individuos) con el control (5 individuos). En cada uno se calculó el PS. Para el ejemplo, la variable de confusión es solo una: número de plaquetas (n° /mCL), amarillos $> 10^5$, azules $\leq 10^5$. Para ajustar esta covariable se aparean individuos de cada grupo con valores similares de PS. La muestra original de 9 se reduce a 2 pares (4 individuos).

La muestra, ajustada según el recuento de plaquetas, ha quedado significativamente reducida en su número.

Figura 5. Inversa de la probabilidad ponderada (IPTW)



En la parte superior se ilustra la condición de la Figura 4. Se pretende ajustar la variable recuento de plaquetas (que para este ejemplo teórico es nuevamente el único confundidor existente), dicotomizada en > 10⁵ (amarillo) y ≤ 10⁵ (azul).

En la parte inferior se ilustra la población teórica luego de aplicar la metodología de IPTW. Con > 10⁵ la probabilidad de intervención, igual en todos los individuos, es 3/4 = 0,75 ; con ≤ 10⁵ la probabilidad de intervención, también igual en todos, es 1/5 = 0,20. Esas probabilidades son el PS, que en este caso teórico, surgen de un simple cálculo matemático.

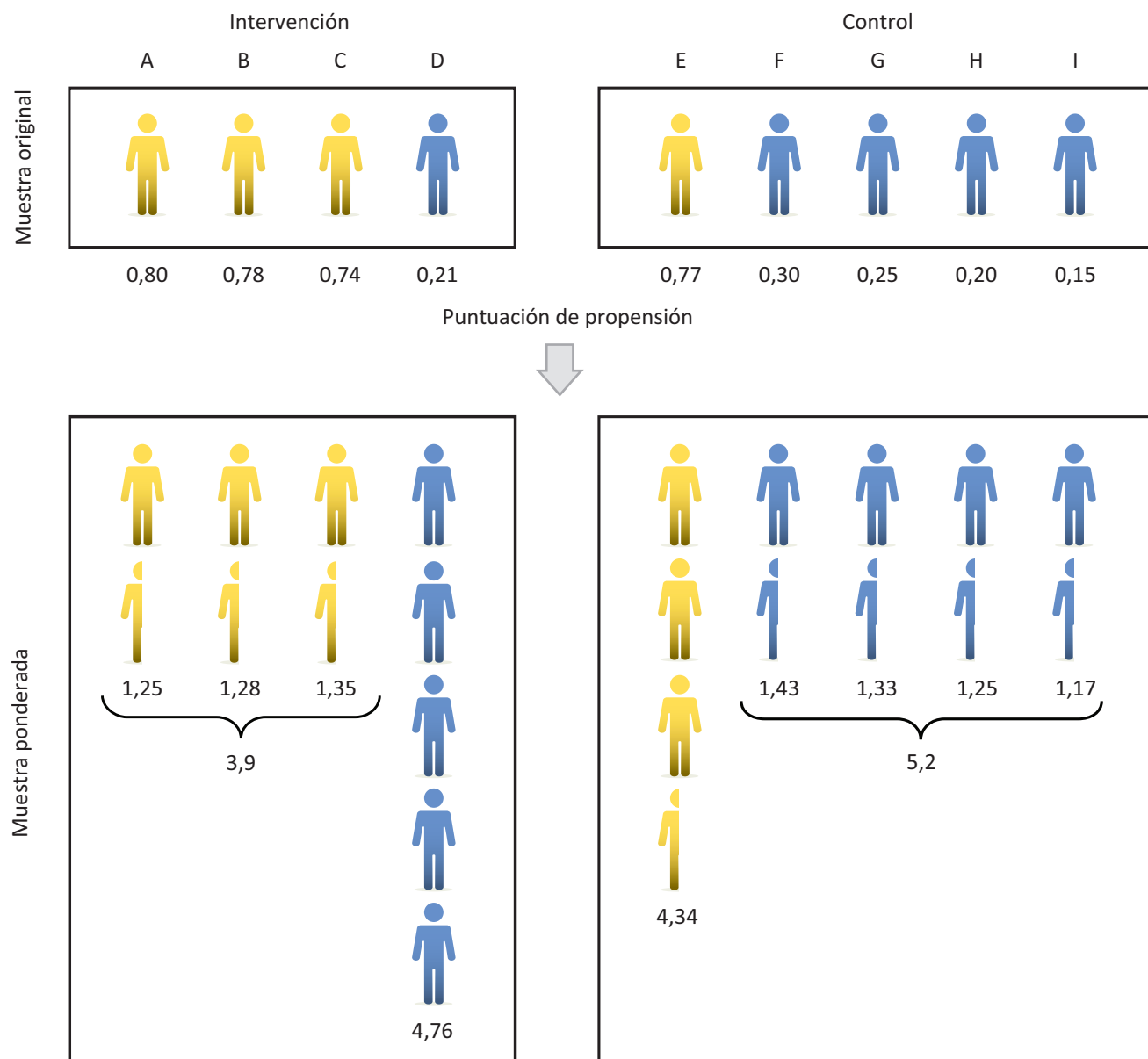
Las probabilidad de recibir control es 1 - PS (1 - 0,20 = 0,80).

Se describe el cálculo correspondiente para cada grupo.

Obsérvese que al final del procedimiento el número de individuos con > 10⁵ plaquetas (amarillos) quedó igualado, 4 con tratamiento y 4 con control; el número de individuos con ≤ 10⁵ (azul) también quedó igualado, en este caso en 5.

Se ha generado una pseudo población con mayor número de casos que la original, pero conservado cada uno de ellos las características basales.

En la tabla adjunta se resume el resultado final y su comparación con el original.

Figura 6. Inversa de la probabilidad ponderada (IPTW)

A diferencia del ejemplo anterior el PS difiere en cada individuo ya que están condicionadas no solo por el recuento de plaquetas sino por múltiples variables de confusión. Su cálculo resulta de un análisis de regresión logística de las covariables con el tratamiento recibido. Aplicando igual metodología que en la Figura 5, pero empleando el PS calculado, el ajuste reduce las diferencias, aunque no es matemáticamente exacto.

Estimando el grado de ajuste

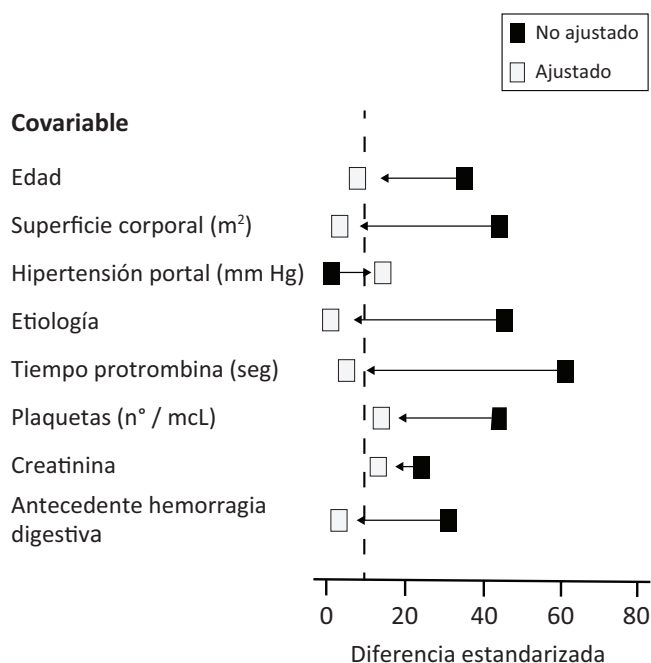
La precisión del ajuste es un aspecto crítico cuando se aplica el método de IPTW para incluir la totalidad de los individuos en estudio, los cuales seguramente presentarán diferencias sustanciales en la prevalencia de múltiples confundidores.

Para evaluar el grado de ajuste alcanzado, se emplea usualmente la diferencia absoluta estandarizada (DAE), es decir la diferencia medida en unidades de desvío estándar,

existente para cada una de las variables de confusión luego del ajuste por IPTW. En general se acepta que el margen que asegura un ajuste adecuado es una diferencia inferior a 0,10, aunque en ocasiones se extiende a $< 0,20$, lo que de alguna manera quita consistencia a las conclusiones del estudio.

La Figura 7 grafica la DAE antes y después del ajuste, mostrando el desbalance previo y el éxito del procedimiento.

Figura 7. Evaluación del ajuste por IPTW en el ensayo teórico



Diferencia estandarizada de las covariables seleccionadas para el ajuste, entre los grupos con y sin tratamiento, antes y después del ajuste por IPTW. El valor considerado como adecuado es 0,10. El ajuste permitió un buen balance de los confundidores. Obsérvese que la diferencia previa al ajuste (cuadros negros) es muy superior a alcanzado post ajuste (cuadros blancos).

PS, IPWT y análisis derivados

En la bibliografía médica actual no es infrecuente que en los estudios observacionales figuren los términos "*Propensity Score*" y "Población IPWT", por lo cual es necesario conocer la base conceptual de estas metodologías estadísticas.

Con ambos procedimientos se trata de disponer de dos poblaciones que, en principio, solo difieran en cuanto a si han recibido o no la intervención objetivo del estudio, ya que el resto de las variables condicionantes de la evolución han sido "igualadas".

Desde aquí, sería posible continuar con el análisis estadístico confeccionando la **curva ajustada de Kaplan y Mayer** y estimar si existe diferencia significativa entre los grupos, empleando el log Rank Test o el Modelo de Cox.

Siguiendo con el ejemplo teórico de este escrito, sería factible graficar la relación entre los *odds ratio* del *outcome* (nuevamente mortalidad hepática o hemorragia digestiva no mortal) con diferentes covariables expresadas como **dato continuo**, por ejemplo el recuento de plaquetas, la

edad o el tiempo de protrombina, considerando de este modo exactamente el **punto de corte** determinante del pronóstico.

Un ensayo observacional realizado con la finalidad de evaluar una intervención o estimar el valor de un índice pronóstico siempre constituye un desafío metodológico. La clave es el procedimiento de ajuste de covariables para hallar el valor real de la asociación motivo de la investigación.

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

Habitualmente, un ensayo clínico tiene por finalidad predecir, a partir de un conjunto de datos disponibles (variables independientes), otro (variable dependiente, punto final o *outcome*), temporalmente presente (estudio diagnóstico) o que aparecerá algún tiempo después (estudio pronóstico). El cualquier caso, el *outcome* queda condicionado a una multiplicidad de esos condicionantes, estadísticamente definidos como variables independientes.

Ahora bien, a partir del resultado se podrá generar un score, como por ejemplo la probabilidad de hemorragia digestiva por várices esofágicas o la mortalidad en la insuficiencia hepática crónica.

En otros casos interesa conocer solo el efecto de una de esas variables independientes.

Ocurre entonces que, si la variable que es motivo de estudio está asociada a otras (variables confundidoras), podría surgir el siguiente interrogante: "¿la evolución favorable es debida a la efectividad del tratamiento en estudio o se debe a que ese tratamiento fue indicado en individuos de menor riesgo?" En estos casos se requiere aislar la asociación de interés de los confundidores, condición referida metodológicamente como ajuste. Los estudios aleatorizados constituyen la opción ideal, ya que al asignar en forma aleatoria el tratamiento, los confundidores quedarán balanceados entre los grupos en observación.

En los estudios observacionales esta problemática es el aspecto central y el mayor condicionante del resultado, sea el objetivo de la investigación la búsqueda de un indicador pronóstico o el efecto de una intervención. La metodología aplicada puede ser, entre otras, el análisis multivariado, el *propensity score* o el inverso de la probabilidad ponderada. Dicha complejidad estadística afecta no solo el diseño sino también la interpretación de la información bibliográfica.

En los ensayos observacionales la información nece-

saría o *input*, proviene de la práctica médica habitual, registrada en la historia clínica o en una base de datos, en ocasiones internacionales y de gran volumen (*Big Data*). De tal modo, al no modificar el accionar médico cotidiano, es claramente un reflejo cabal del “mundo real”, del accionar asistencial y no una condición experimental, como ocurre en los estudios aleatorizados. Ello mismo es determinante, además, del menor costo y factibilidad para concretar la investigación.

El espectro de la validez metodológica de un ensayo observacional es amplio, donde en un extremo está el registro de datos históricos de la historia clínica, frecuentemente con información faltante u obtenida sin una sistemática predefinida. El otro extremo corresponde a un diseño con valor metodológico superior, el cual requiere un protocolo *ad hoc*, con una precisa metodología estadística y de otros condicionantes imprescindibles, como precisar las condiciones de inclusión y exclusión y la definición *a priori* de los puntos finales primarios y secundarios, entre otros parámetros.

La opción ensayos aleatorizados vs. observacionales es falsa: ambas metodologías se complementan en el avance del conocimiento médico.

Tal vez, lo que sí merece una consideración especial, es la afirmación del físico Richard Feynman cuando concluye: “es mucho más interesante vivir con cierta incertidumbre, que vivir con respuestas que pueden estar equivocadas”. (*It is much more interesting to live with uncertainty than to live with answers that might be wrong*).

Propiedad intelectual. Los autores declaran que los datos presentes en el manuscrito son originales y se realizaron en las instituciones a las cuales pertenecen.

Financiamiento. Los autores declaran que no hubo fuentes de financiación externas.

Conflicto de interés. Los autores declaran no tener conflictos de interés en relación con este artículo.

Aviso de derechos de autor



© 2025 Acta Gastroenterológica Latinoamericana. Este es un artículo de acceso abierto publicado bajo los términos de la Licencia Creative Commons Attribution (CC BY-NC-SA 4.0), la cual permite el uso, la distribución y la reproducción de forma no comercial, siempre que se cite al autor y la fuente original.

Cite este artículo como: Ciocca M, Mastai R y Cagide A. El estudio observacional. Acta Gastroenterol Latinoam. 2025;55(2):70-78. <https://doi.org/10.52787/agl.v55i2.489>

The Observational Study

Mirta Ciocca¹  · Ricardo Mastai¹  · Arturo Cagide² 

¹ German Hospital of Buenos Aires.

² Italian Hospital of Buenos Aires.

City of Buenos Aires, Argentina.

Acta Gastroenterol Latinoam 2025;55(2):79-87

Received: 05/04/2025 / Accepted: 06/02/2025 / Published online: 30/06/2025 / <https://doi.org/10.52787/agl.v55i2.489>

A Fictional Story

At a center specializing in liver diseases, a new drug ("Tx. X") has been used to treat portal hypertension. The professionals are convinced that the drug reduces gastrointestinal bleeding associated with this syndrome.

A statistical analysis shows that with Tx. X, there was a significant decrease in that complication, with $p < 0.05$.

But a question arises: Did only patients with better clinical conditions and favorable prognosis, such as younger age, lower degree of portal hypertension, or an absence of thrombocytopenia, receive the treatment? If so, is the $p < 0.05$ due to the treatment or because the intervention acted as a "selector" of a lower-risk population?

Observational Design

Figure 1 illustrates a fictional observational study, in which the exposure variable (the research objective) could be associated with the evolution variable (*the end point or "outcome"*) or with a certain number of *confounders* (or covariates) that can affect that association.

For instance, as previously mentioned, if an intervention that can favorably modify the outcome –in this case, liver-related mortality or non-fatal gastrointestinal bleeding– is preferentially indicated for patients with a higher platelet counts or less severe disease, the benefit could be due to the treatment or the fact that the procedure "selected" individuals with a lower bleeding risk. Note that this detailed condition requires the association of confounders with the intervention.

Correspondence: Mirta Ciocca
Email: cioccamirta@gmail.com

In other words, any **imbalance** in the prevalence of confounders between comparison groups can distort the target association. Thus, when only the relationship between the intervention and the outcome is considered using **univariate** or **unadjusted** analysis, the resulting *p* value may have no clinical significance with regard to the treatment effect.

This situation does not affect randomized (experimental) trials because the decision to intervene is based on chance rather than medical necessity, ensuring balanced distribution of confounders in both intervention and control groups. However, randomization can occasionally fail, especially in small samples.

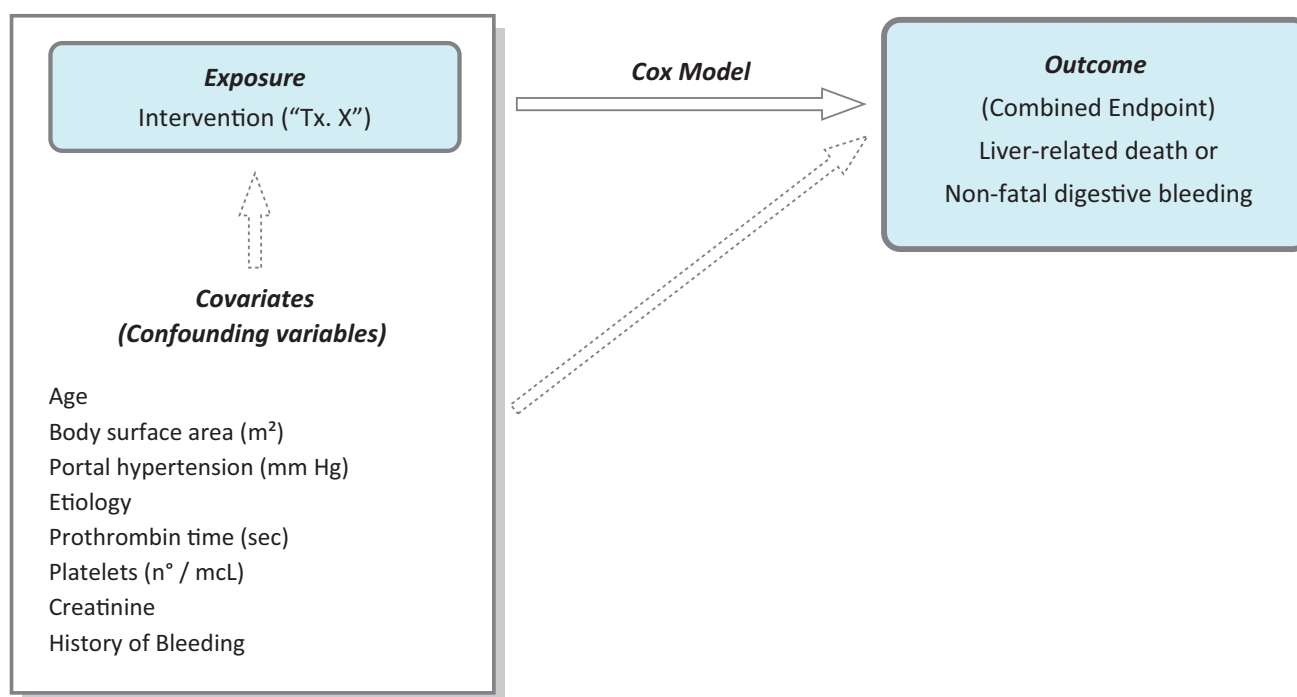
The usual procedure to address this issue in a follow-up trial is the **multivariate method of logistic re-**

gression or Cox proportional hazards model (hereafter, “Cox model”).

With this methodology, the exposure variable and the confounders (selected for their isolated statistical association or the bibliographic contribution **-independent variables-**) are analyzed together (Figure 1). This procedure is usually referred to as **adjustment**. In multivariate analysis, if the exposure variable is statistically significant in relation to the outcome variable (**dependent variable**), it is concluded that this association is not conditional on confounders.

However, multivariate analysis requires a specific ratio of independent variables to outcome variable prevalence/incidence, which limits the applicability of this methodology under certain conditions.

Figure 1. Multivariate Study



The exposure variable (intervention) and its possible association with the outcome variable are shown (shaded areas, continuous line arrow). If confounding variables are associated with both the exposure and the outcome (dashed arrows), the exposure/outcome association could be attenuated or diluted. The multivariate analysis of the set (shaded box) estimates the adjusted association of exposure with outcome, independent of the confounders.

Propensity Score (PS)

A second statistical methodology is balancing the confounders so that they are equal between the intervention and control groups.

Figure 2 illustrates the procedure. First, the association between confounders (now independent variables)

and the exposure (now dependent variable) is estimated statistically using multivariate analysis (logistic regression in this case).

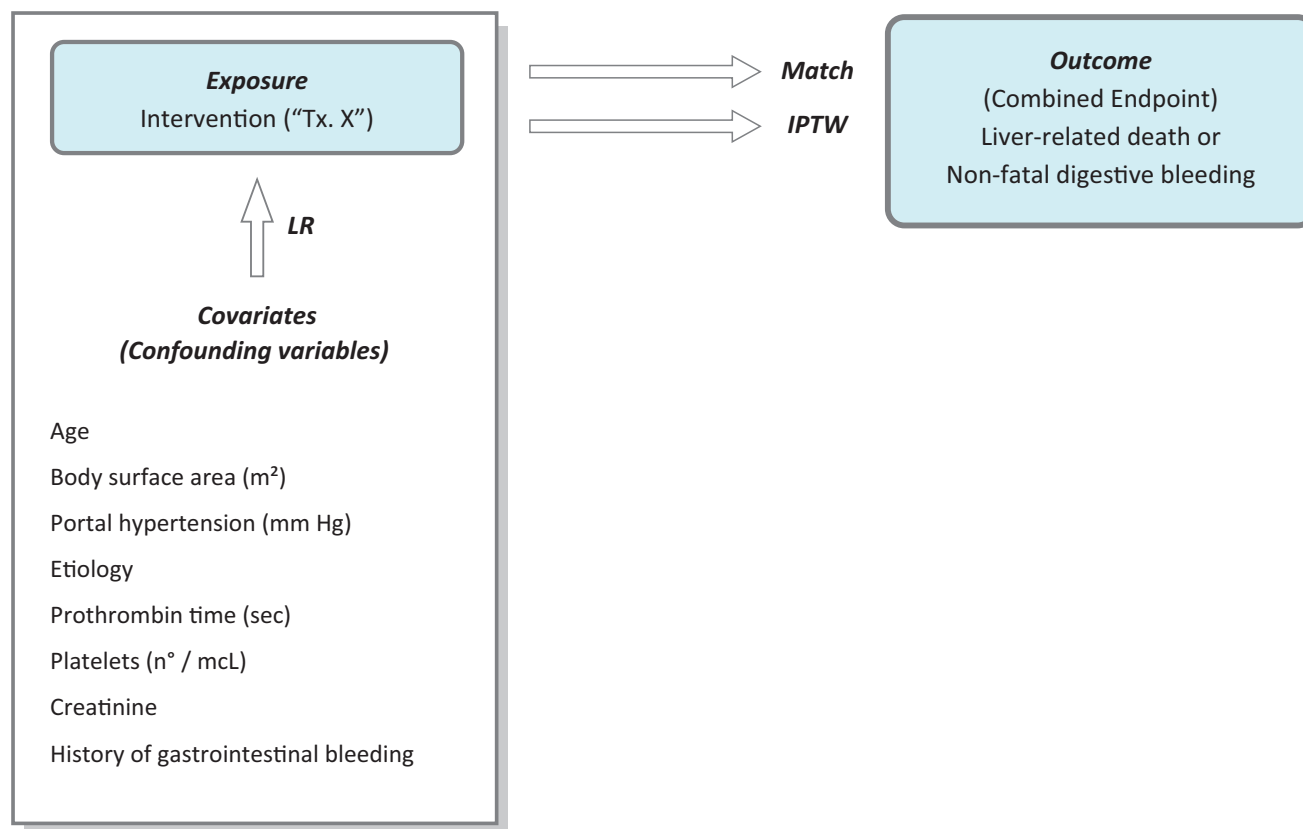
The **Propensity Score (PS)** which is the probability of being exposed to the intervention independently of the confounders, is derived from this analysis. Individ-

uals with similar PS should have an equal probability of receiving the treatment under investigation, whether or not they actually received it.

An advantage of PS over multivariate analysis is

that the number of independent variables is not limited by the prevalence of the outcome, since the intervention group, will always have a sufficient number of observations.

Figure 2. Propensity Score and Derived Analyses



The final objective of the study, as in Figure 1, is to estimate the possible association between exposure (intervention, "Tx. X") with the endpoint (outcome). Previously, the association of confounders with exposure is analyzed, and the Propensity Score (PS) is calculated.

From there, the PS allows, by applying different statistical methodologies, such as matching or the inverse probability of treatment weighting (IPTW), to estimate the independent association of the exposure with the outcome.

LR = Logistic Regression.

Exposure and Outcome Variables

There is some debate about which variables should be included in the PS calculation. In general, they should include all those that the investigator considers to condition a given treatment or intervention. In principle, outcome variables should also be included.

However, as in any multivariate analysis, only known and available independent variables are considered, which can result in flawed PS prediction of exposure to the treatment or intervention.

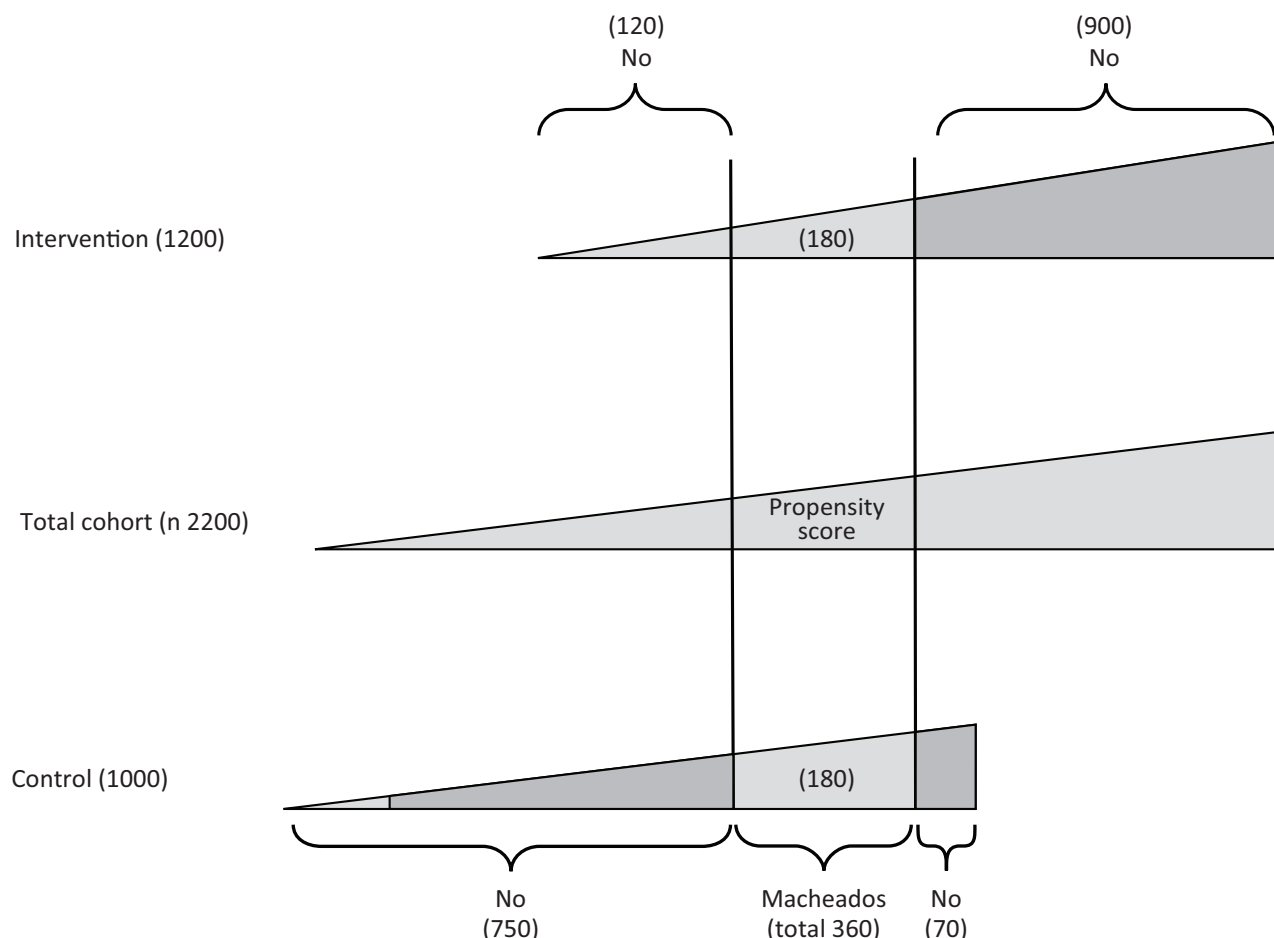
The ROC curve can be applied to calculate the accuracy of the model in the calculation of PS. Although there

are discrepancies in the value considered adequate, most authors consider 0.80 acceptable.

PS, Covariate Adjustment, and Outcome Estimation

Matching

Following the above methodology, each individual will be characterized by a certain PS according to his or her baseline characteristics or confounders. Some individuals in the intervention group will have a similar PS to those in the control group, so that it is possible to **match** patients in both groups according to their PS (Figure 3).

Figure 3. Matching

Theoretical example of a study of 2200 patients, 1200 with the intervention and 1000 controls.

The PS (propensity score) of the whole sample is plotted in an increasing value scale: center = full cohort, top = treated individuals, bottom = controls. Each treated individual is matched with a control with equal or very similar PS. A total of 180 pairs are integrated; the sample is reduced to 16% of the original number (360/2200).

In the matched sample, confounders from both intervention and control groups are **balanced**.

However, a significant number of individuals from both groups will be excluded because the corresponding pair is not available. The number of excluded individuals is directly related to the degree of confounder imbalance between the two groups in the original study sample. Thus, the conclusions of the trial and its application are limited to the matched sample and cannot be generalized to the entire population.

Inverse Probability of Treatment Weighting (IPTW)

Unlike matching, **IPTW** includes the entire sample under study.

While matching achieves adjustment by reducing the population until the confounders are equalized in the groups to be compared, IPTW achieves this objective

by increasing the population with individuals who have a similar rate of confounders through a **mathematical process**.

Figure 4 shows a theoretical example that compares the intervention group with the control group. In this example platelet count is the only confounder considered, and it is dichotomized into > 105 mcL and ≤ 105 mcL).

There is an obvious imbalance, as there are three individuals in the treated group with platelets > 105 mcL and only one in the control group. The covariate platelets must be adjusted so that the two groups can be compared for a given outcome, for example, liver-related mortality or non-fatal gastrointestinal bleeding. For this purpose the PS in each is estimated.

If the matching strategy is applied, two pairs of 2 patients each, treated and control, sharing similar PS could be integrated: the sample would be limited to only four individuals (Figure 4).

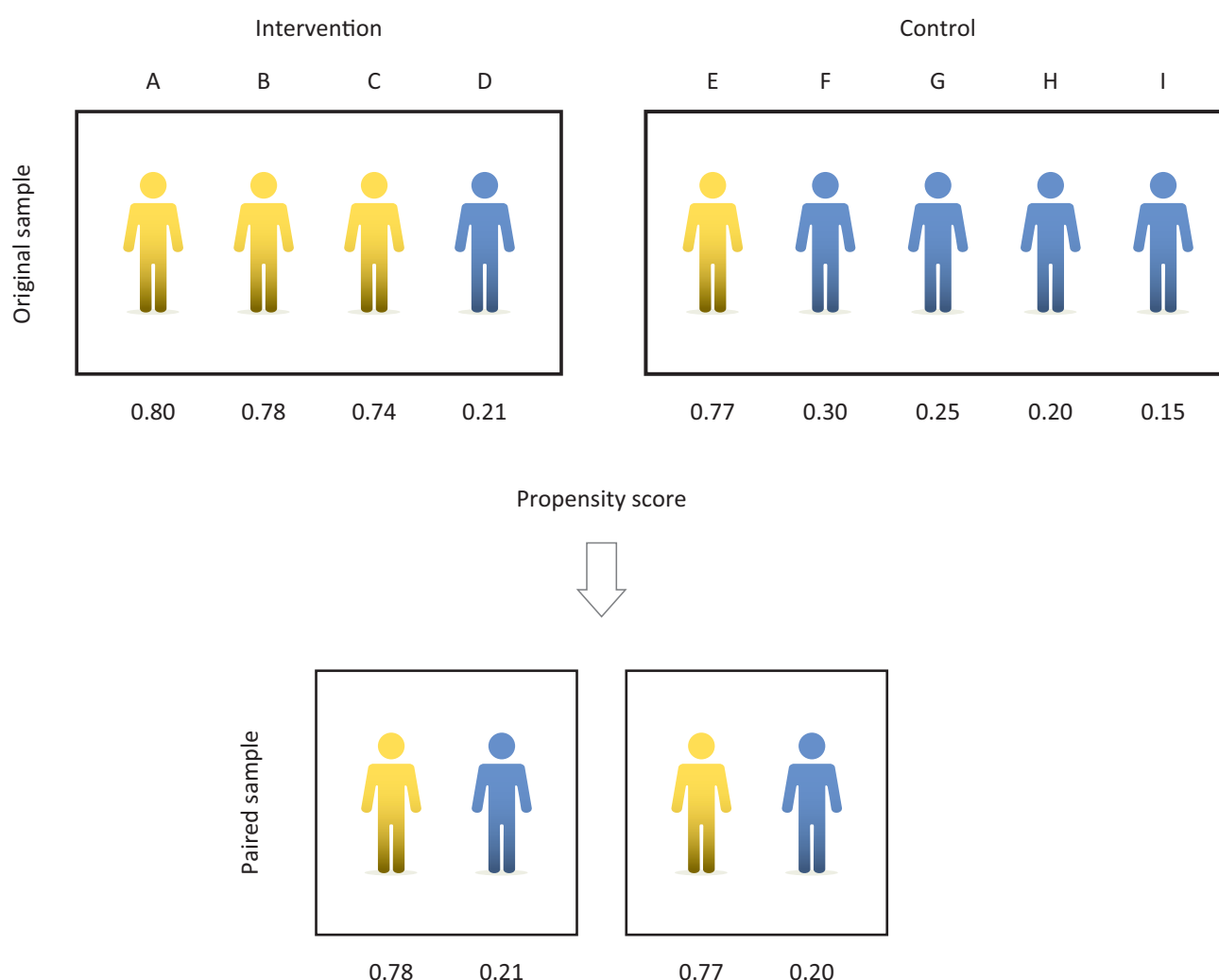
In the same example, if the IPTW is used, the platelet covariate is adjusted by artificially increasing the number of observations, as previously mentioned, using the procedure detailed in Figure 5.

The PS is also used in the calculation, but in this case, by not reducing the number of observations, the result of the investigation can be generalized to a population that is more representative of real-world clinical practice.

However, in a real context, the probability of treatment is conditional to multiple confounders, whose overall effect is represented in the PS, which will differ for each individual (Figure 6). The procedure, similar to the one explained in Figure 5, although it reduces the imbalance, it is not perfect, and some differences persist in the distribution of individuals with thrombocytopenia between the treated and untreated groups, which is due to the effect of the other confounders.

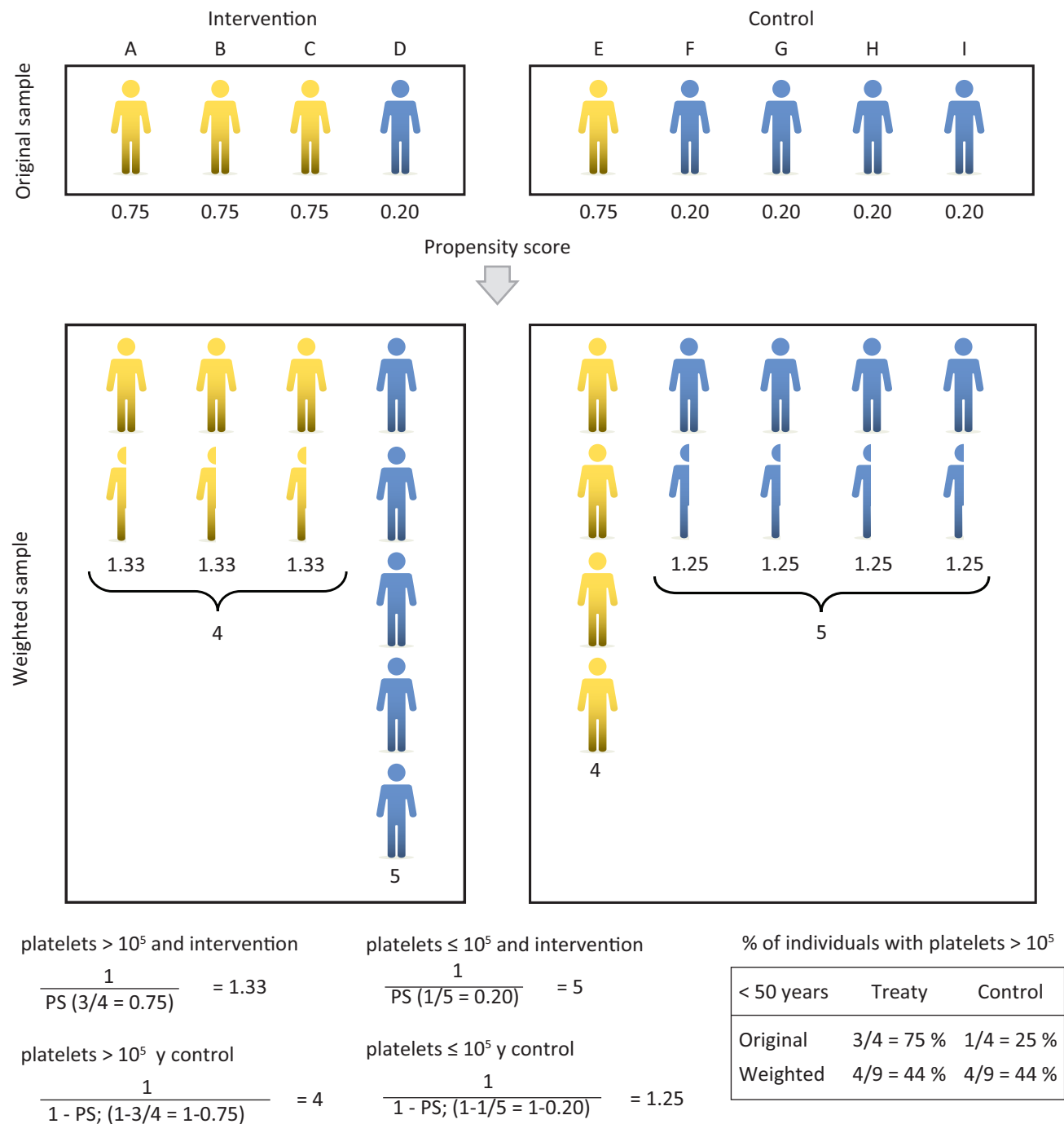
The strategy detailed for platelets should be applied to all available variables or confounders considered when calculating PS.

Figure 4. Matching



Theoretical example of an observational study comparing intervention (4 individuals) with control (5 individuals). In each, the PS was calculated. For the example, the confounding variable is only one: number of platelets (n°/mCL), yellow $> 10^5$, blue $\leq 10^5$. To adjust for this covariate, individuals from each groups with similar PS values are paired. The original sample of 9 is reduced to 2 pairs (4 individuals). The sample adjusted for platelet count has been significantly reduced in number.

Figure 5. Inverse Probability of Treatment Weighting (IPTW)



The top illustrates the scenario of Figure 4. The platelet count variable (the only existing confounder in this example), dichotomized as > 10⁵ (yellow) and ≤ 10⁵ (blue), is intended to be adjusted.

The bottom shows the theoretical population **after IPTW**: with > 10⁵ the probability of the intervention is 3/4 = 0.75; with ≤ 10⁵ the probability of the intervention is 1/5 = 0.20. These probabilities are the PS, obtained by a simple mathematical calculation.

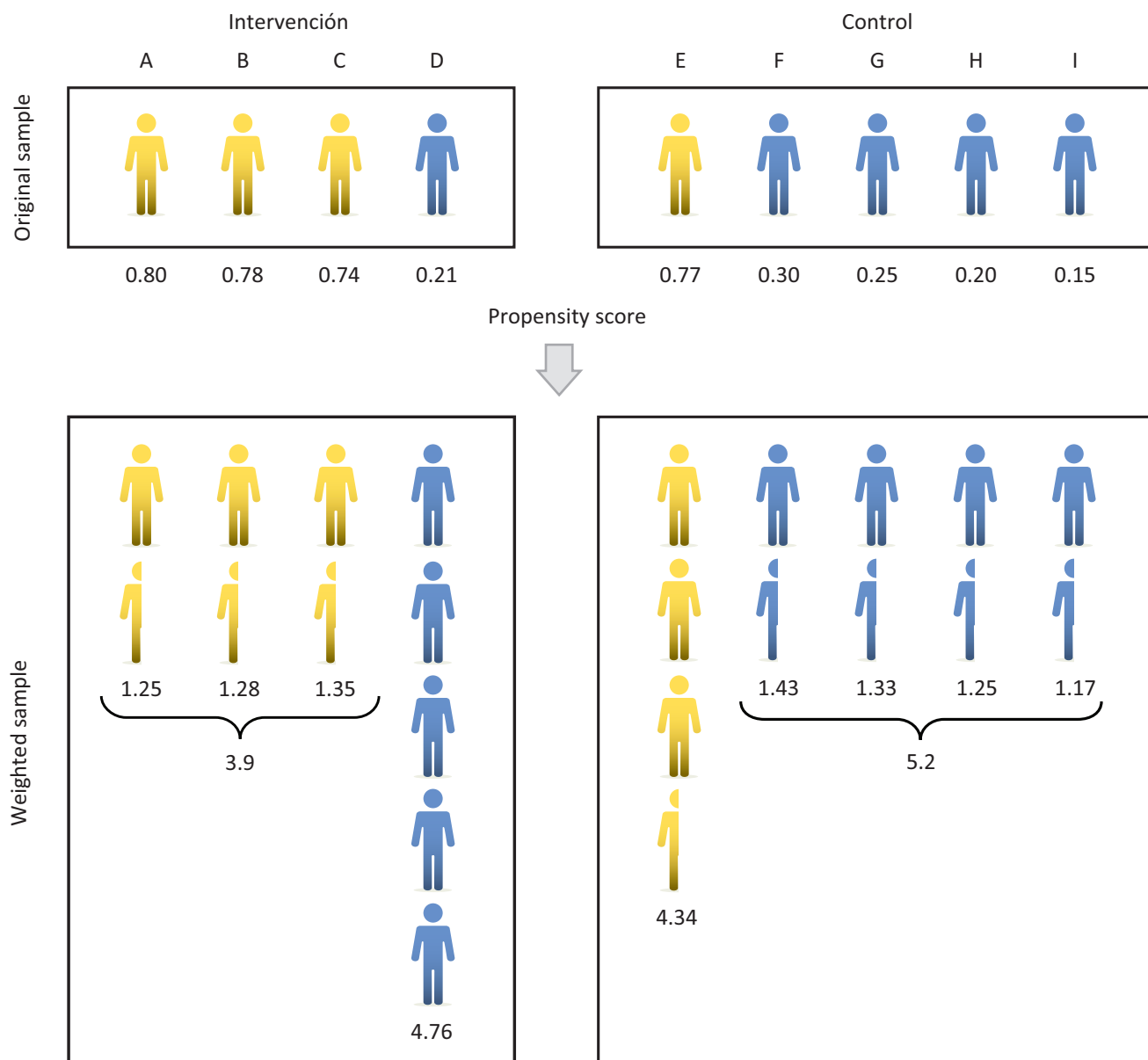
Control probability = 1 - PS (e.g., 1 - 0.20 = 0.80).

The corresponding calculations is described for each group.

At the end of the procedure the number of individuals with > 10⁵ platelets (yellow) was equalized, 4 treated and 4 control. The number of individuals with ≤ 10⁵ (blue) was also equalized, in this case at 5.

A **pseudo-population** was generated with a greater number of cases than the original, but baseline characteristics are preserved.

The attached table summarizes the final results and its comparison with the original.

Figure 6. IPTW (*Realistic Scenario*)

Unlike the previous example, here the PS differs in each individual since they are conditioned not only by the platelet count but also by **multiple confounding variables**. PS is calculated via a logistic regression of the covariates with the treatment received. Using the same IPTW method as Figure 5, the adjustment reduces the differences but it is not mathematically exact.

Estimating the Degree of Adjustment

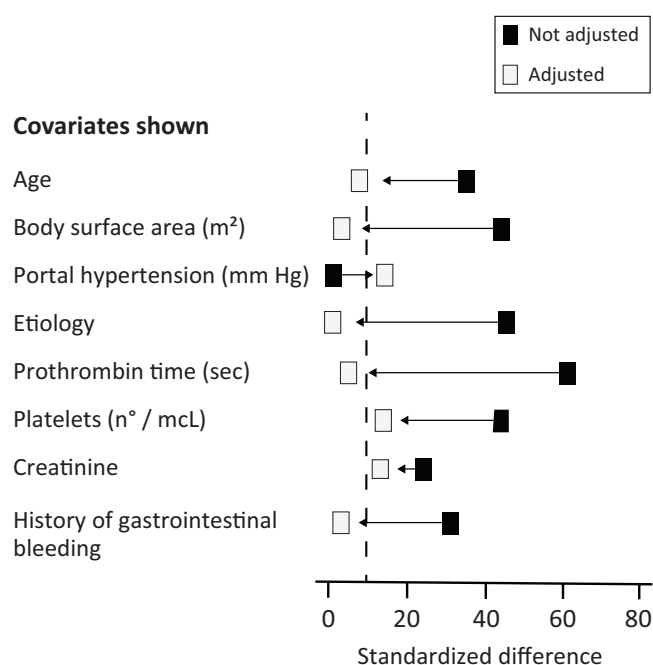
Accurate adjustment is critical when applying the IPTW method to include all the individuals under study, as they will certainly have substantial differences in the prevalence of multiple confounders.

The **standardized absolute difference (SAD)**, or the difference measured in standard deviation units, is typically used to assess the degree of adjustment

achieved for each confounding variable after IPTW adjustment. Generally, it is accepted that an adjustment is adequate if the difference is less than 0.10, although it sometimes extends to less than < 0.20 , this detracts from the consistency of the study's conclusions.

Figure 7 plots the SAD before and after adjustment to show the prior imbalance and the success of the procedure.

Figure 7. Evaluation of IPTW Adjustment in the Theoretical Trial



Standardized difference of the covariates selected for adjustment, between the groups with and without treatment, before and after IPTW adjustment. The value considered adequate is 0.10.

The adjustment allowed a good balance of confounders.

Note that pre-adjustment differences (black boxes) were much higher than post-adjustment (white boxes).

PS, IPTW, and Derived Analyses

The terms “**propensity score**” and “**IPTW population**” are frequently used in observational studies in the current medical literature, so it is necessary to understand the conceptual basis of these statistical methodologies.

The goal of both procedures is to create two populations that differ only in terms of whether or not they have received the intervention targeted by the study. This is because the rest of the conditioning variables that affect evolution have been “equalized.”

From here, one could continue with the statistical analysis by creating an adjusted Kaplan-Meier curve and using the Log-Rank Test or the Cox Model to estimate whether there is a significant difference between the groups.

Using the theoretical example of this paper, one could plot the relationship between the odds ratio of the outcome (e.g., liver-related mortality or non-fatal gastrointestinal bleeding) and different continuous covariates,

such as platelet count, age, or prothrombin time. This would allow one to consider the exact prognostic cut-off point.

An observational trial designed to evaluate an intervention or estimate the value of a prognostic index poses a **methodological challenge**. The key lies in the covariate adjustment procedure, which is used to determine the **true value of the association** under investigation.

The accuracy of PS in estimating the probability of intervention is a determining factor in the IPTW methodology. Once again, unmeasured or unknown confounders are a critical point of the statistical procedure.

Typically, a clinical study seeks to **predict** one variable (the dependent variable, endpoint, or outcome) from a set of independent variables, whether that prediction concerns present (diagnostic) or future (prognostic) outcomes. In any case, the outcome is conditioned by a multitude of these factors, which are statistically defined as independent variables.

However, a score can be generated from the outcome. For example, one could calculate the probability of gastrointestinal bleeding due to esophageal varices or mortality in chronic liver failure.

In other cases, it is of interest to know the effect of only one independent variable. If that variable is associated with others, known as confounding variables, the following question may arise: “Is the favorable outcome due to treatment effectiveness or it is because the treatment was indicated for individuals at lower risk?” In these cases, it’s necessary to isolate the association of interest from the confounders. This process is referred to as an adjustment in methodology. Randomized trials are the ideal option since confounders are balanced between groups when treatment is randomly assigned.

In observational studies, this problem is central and determines the outcome, whether the goal is assessing a prognostic indicator or to study the effect of an intervention. The applied methodology can include multivariate analysis, propensity score or inverse probability weighting. This statistical complexity affects the design and interpretation of bibliographic information.

In observational trials, necessary information comes from routine medical practice and is recorded in clinical histories or databases, sometimes international and large-scale (big data). Since it does not modify daily medical practice, it is clearly a true reflection of health care practice in the “real-world” and not an experimental condition, as occurs in randomized trials. This also explains the lower cost and feasibility of conducting the research.

The spectrum of methodological validity of an ob-

servational trial is broad. At one extreme is the recording of historical data from clinical histories, which often contains missing information or was obtained without a predefined, systematic approach. At the other extreme is a design with superior methodological value. This design requires an *ad hoc* protocol with a precise statistical methodology and other essential conditions. These conditions include specifying the inclusion and exclusion criteria, and defining the primary and secondary endpoints *a priori*, among other parameters.

The randomized versus observational trial option is false: both are complementary methodologies that advance medical knowledge.

What does deserve special consideration is the apt statement by physicist Richard Feynman: ***“It is much more interesting to live with uncertainty than to live with answers that might be wrong.”***

Intellectual property. The authors declare that the data in the article are original and were carried out at their institutions.

Funding. The authors declare that there were no external sources of funding.

Conflict of interest. The authors declare that they have no competing interests related to this article.

Copyright



© 2025 Acta Gastroenterológica latinoamericana. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY-NC-SA 4.0) license, which allows non-commercial use, distribution, and reproduction, provided the original author and source are acknowledged.

Cite this article as: Ciocca M, Mastai R y Cagide A. The Observational Study. Acta Gastroenterol Latinoam. 2025;55(2):79-87. <https://doi.org/10.52787/agl.v55i2.489>