

Propensity Score

Algunes idees per al seu ús en la Recerca Clínica

Unitat d'Estadística I Bioinformàtica

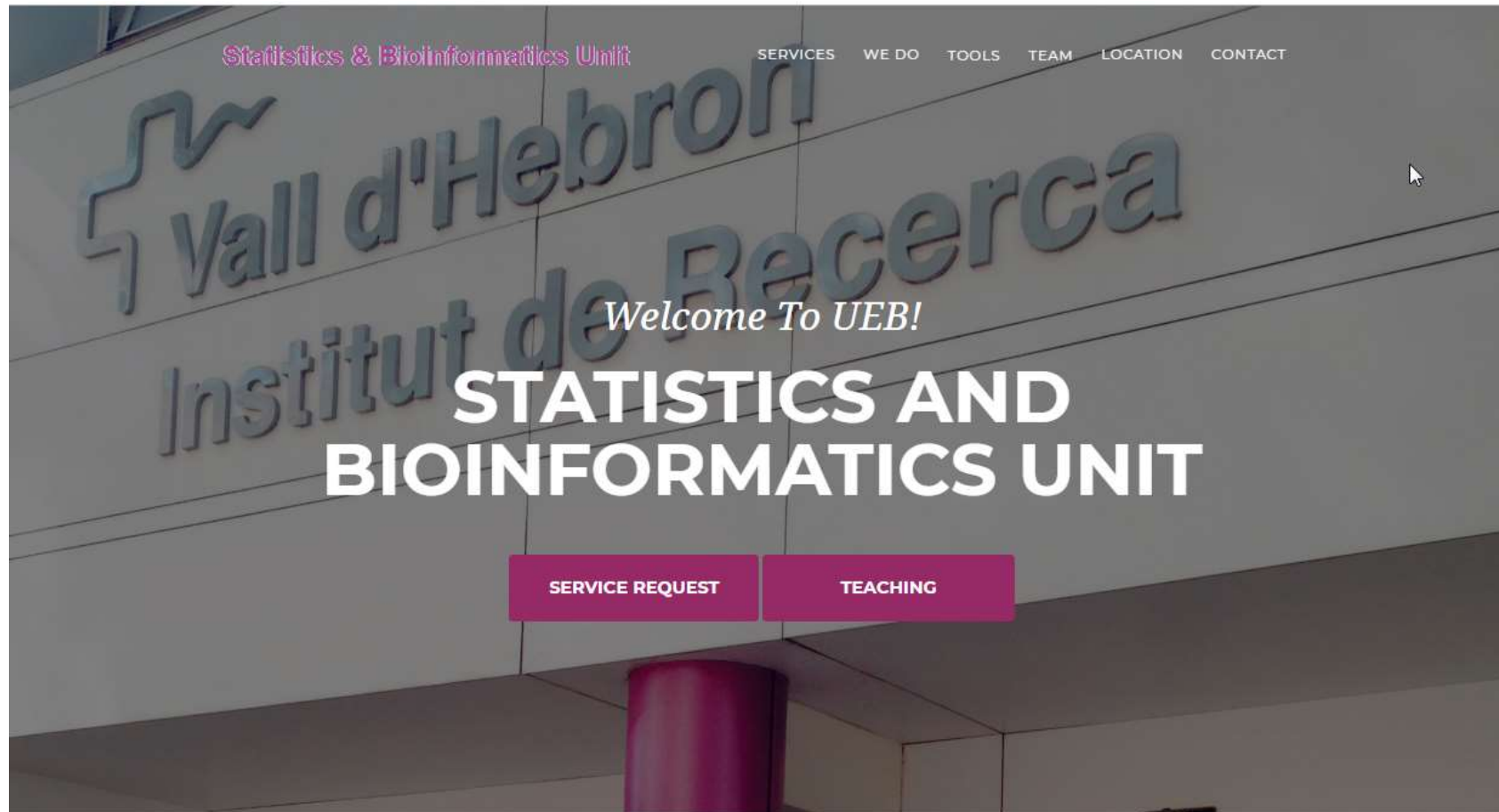
29 d'Abril de 2022



**PINDOLLES
ESTADÍSTIQUES
UEB-VHIR**



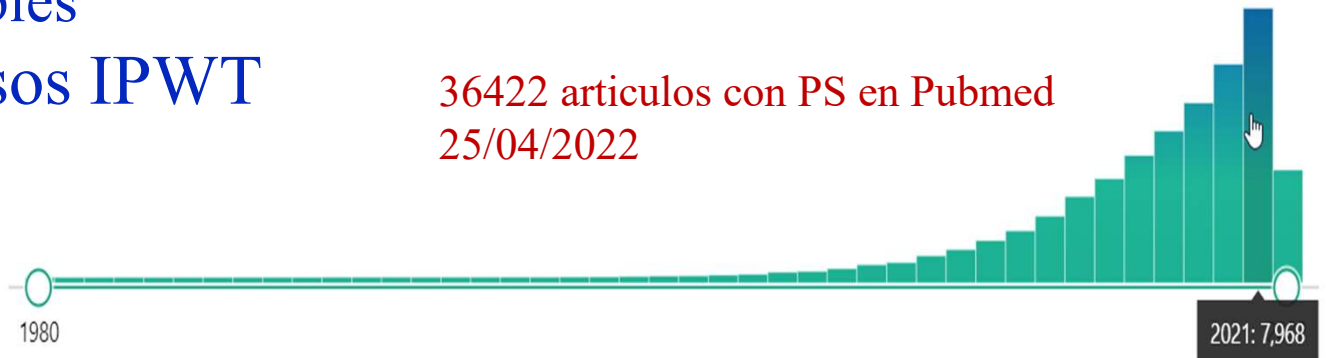
ueb.vhir.org



Contenido de la Sesión

- Introducción
- Riesgo Causal. Contrafactual
- Medidas de efecto
- Ensayo Clínico vs Estudio Observacional
- Metodos de Propensity Score
- Calculo de Propensity Score
- Matching
- Equilibrio de Covariables
- Uso de pesos Inversos IPWT
- Resumen

36422 articulos con PS en Pubmed
25/04/2022



Introducción



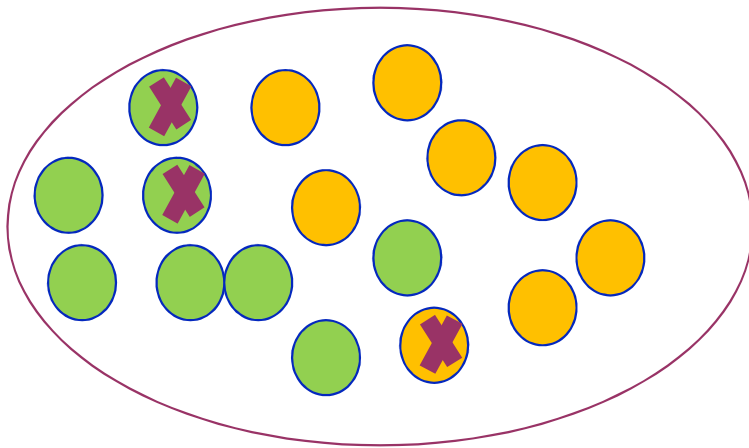
- Patrón de oro para estimar el efecto de los tratamientos
- Asignación del tratamiento al azar.
- Habitualmente estos efectos se interpretan como causales



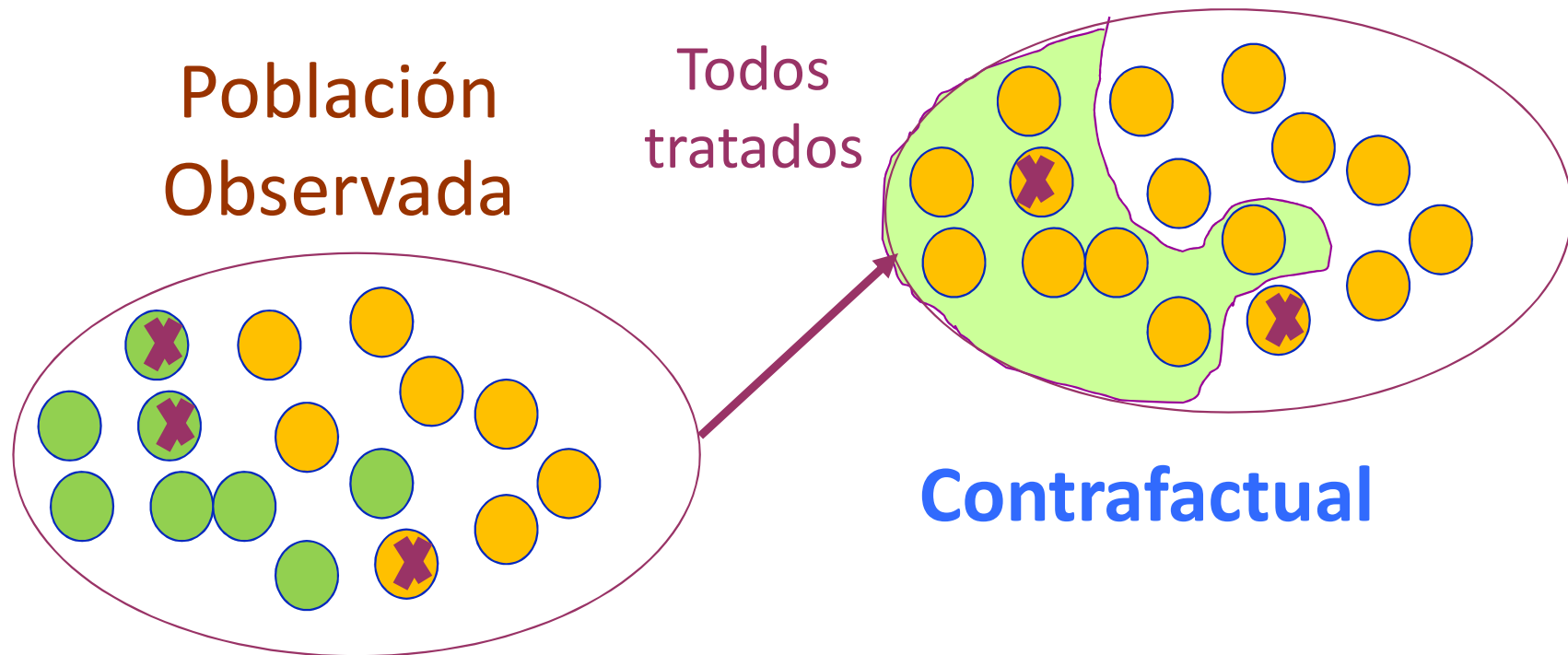
- Información de los tratamientos a largo plazo
- Diferentes estrategias terapéuticas para cada individuo
- Sesgos por indicación y selección

Riesgo Causal

Población
Observada

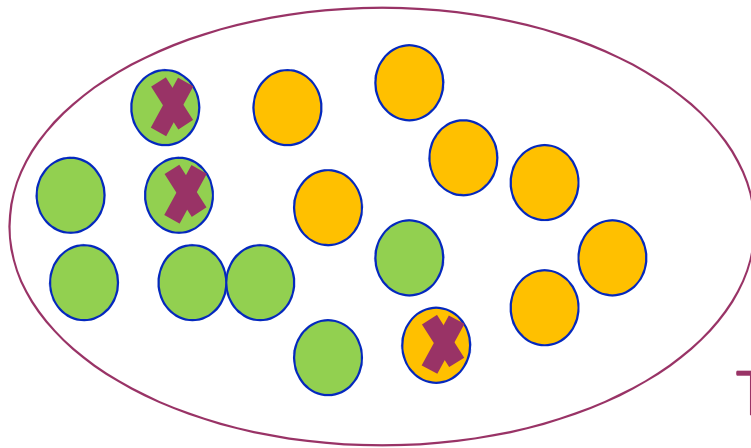


Riesgo Causal



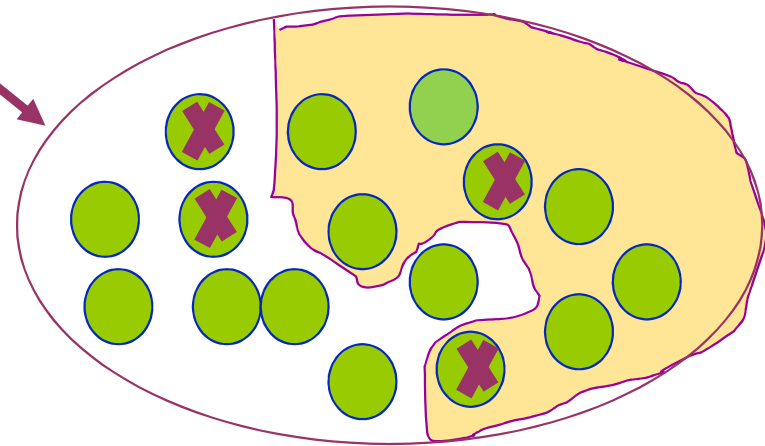
Riesgo Causal

Población
Observada

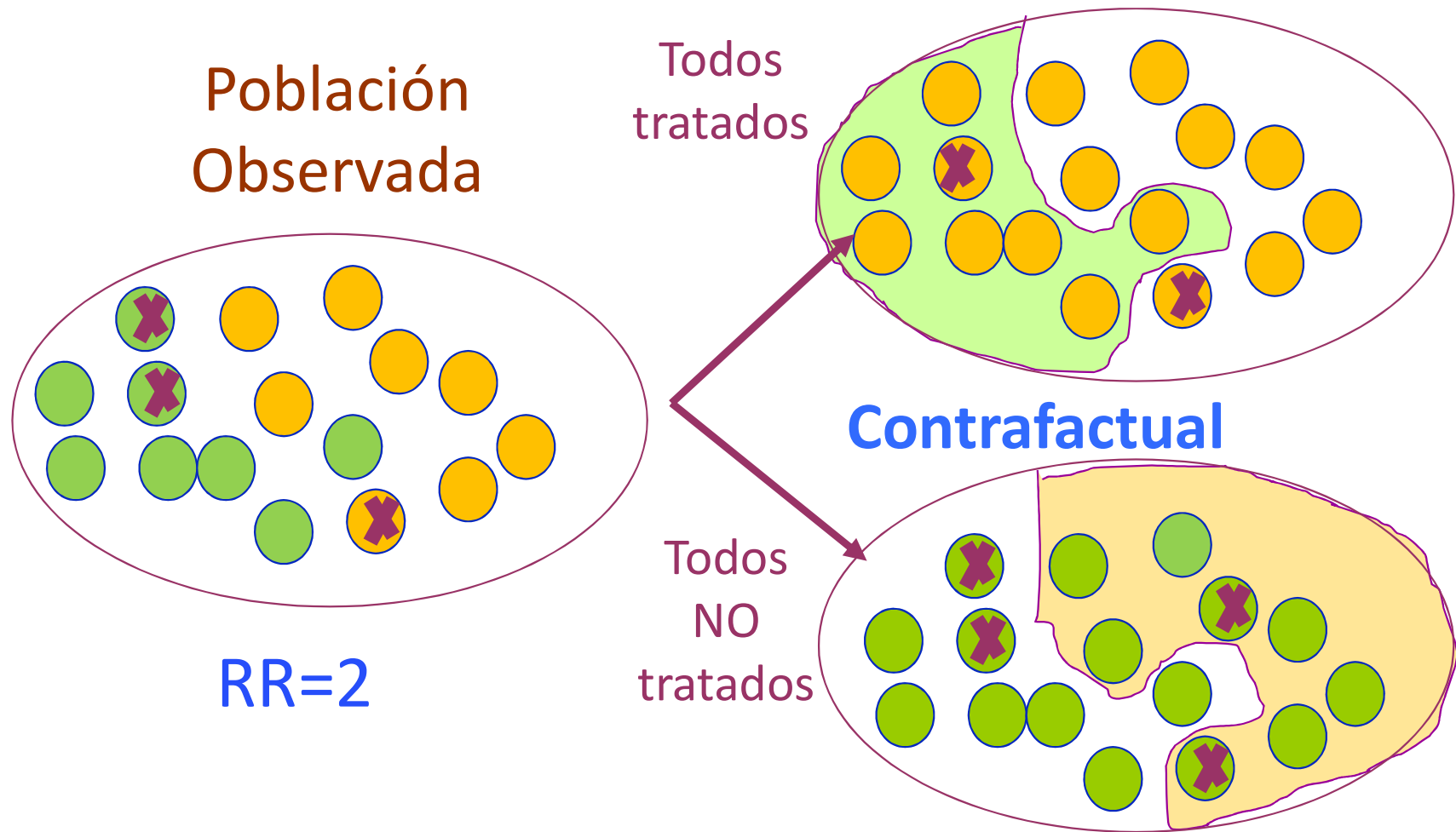


Contrafactual

Todos
NO
tratado
s



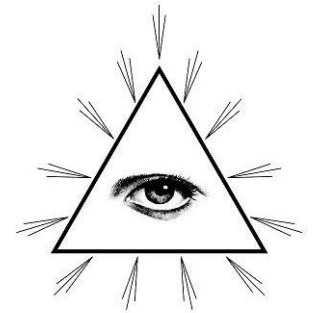
Riesgo Causal



Riesgo Relativo Causal

Exceso o defecto de riesgo que se hubiera observado si se pudiera repetir el experimento **tratando a todos** primero y después **no tratándolos**

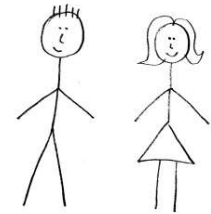
Sea $Y_{a=i}$ la variable **contrafactual** que tiene como resultado lo que ocurriría si todos estuvieran tratados o no



$$RR_{\text{causal}} = P(Y_{a=1}=1) / P(Y_{a=0}=1)$$

Asociación (lo que observamos)

Exceso o defecto de riesgo que se observa entre los tratados y no tratados del estudio.



$$RR_{\text{asociación}} = P(Y_{a=1} | A=1) / P(Y_{a=0} | A=0)$$

Este riesgo no tiene porque ser causal

Efectos potenciales

Efecto Individual (Individual Treatment Effect)

Diferencia del resultado si se trata o no a un individuo

$$Y_{a=1}^i - Y_{a=0}^i$$

Efecto tratamiento medio (Average Treatment Effect)

Diferencia media del resultado si se trata a todos los individuos o no

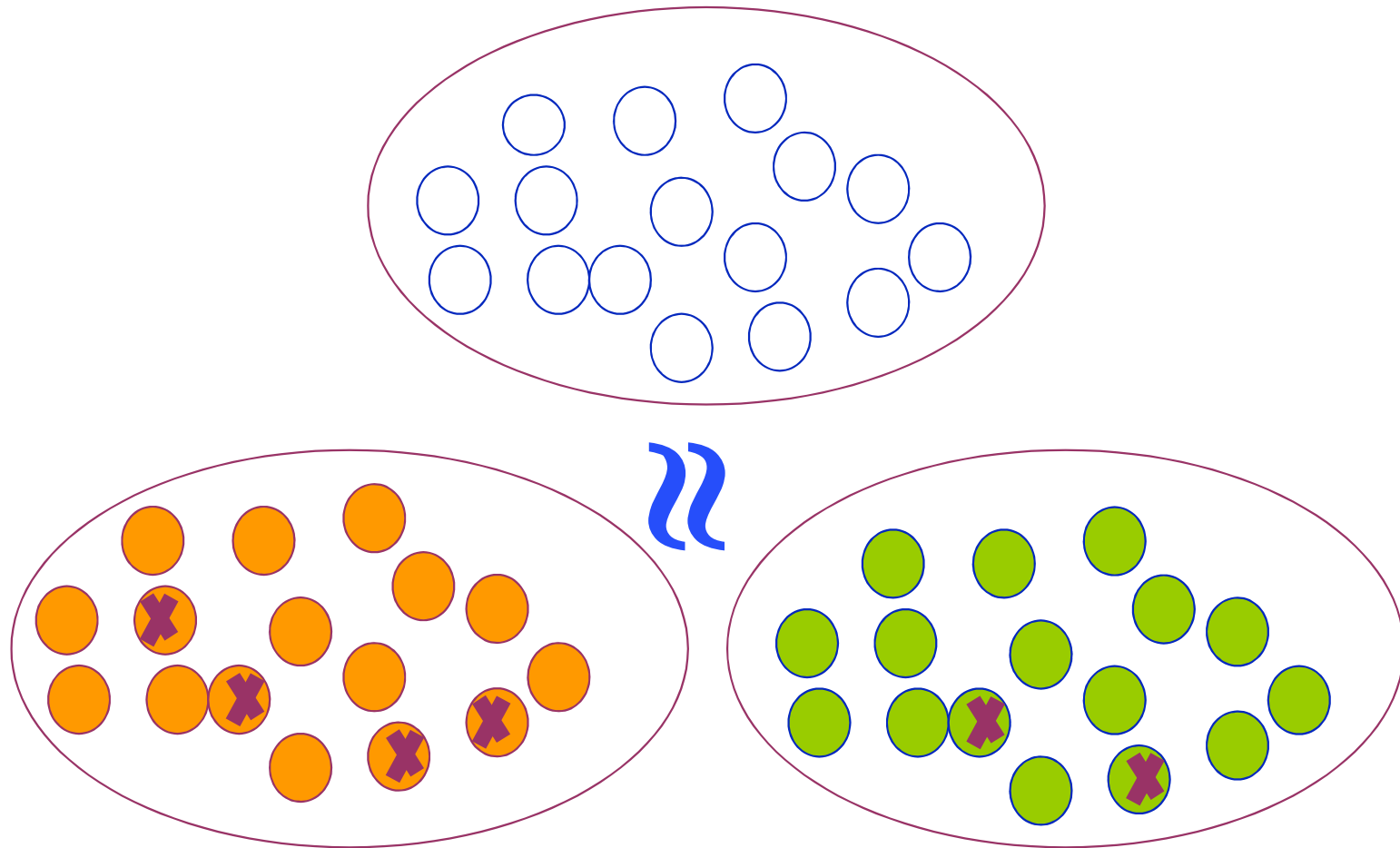
$$E[Y_{a=1}^i - Y_{a=0}^i]$$

Efecto tratamiento medio en tratados (Average Treatment effect in Treated)

Diferencia media del resultado en los tratados si no se trataran

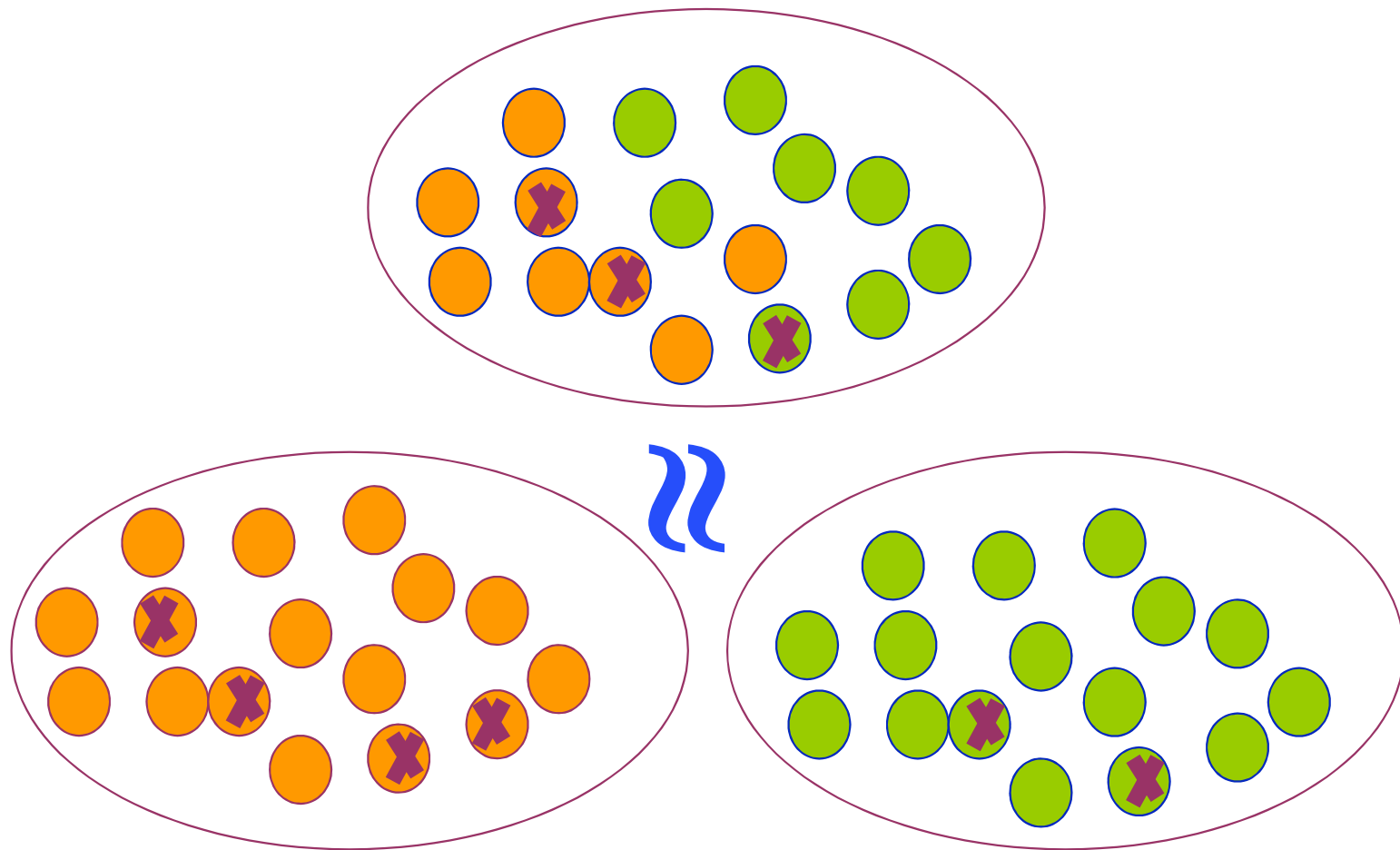
$$E[Y_{a=1}^i - Y_{a=0}^i | A=1]$$

Ensayo Clínico



Hiperpoblación contrafactual

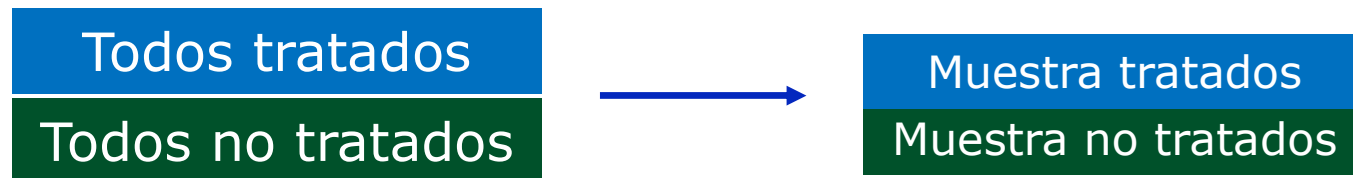
Ensayo Clínico



Hiperpoblación contrafactual

Idea Intuitiva

- **Ensayo Clínico**
- Los sujetos se asignan al azar
- Los sujetos son intercambiables y si se trataran a los no tratados los resultados serian los mismos



$$RR_{\text{causal}} = P(Y=1/A=1) / P(Y=1/A=0) = RR_{\text{observado}}$$

- Pérdidas seguimiento
- Problemas con el tratamiento. Cumplimiento
- Etc.

Intercambiabilidad

Se sabe que pasa cuando

Los realmente expuestos están expuestos

$$E(Y_{a=1} | A=1)$$

Los realmente no expuestos están no expuestos

$$E(Y_{a=0} | A=0)$$

Para obtener $E(Y_{a=1})$ y $E(Y_{a=0})$ necesitamos saber

Los realmente expuestos están no expuestos

$$E(Y_{a=0} | A=1)$$

Los realmente no expuestos están expuestos

$$E(Y_{a=1} | A=0)$$

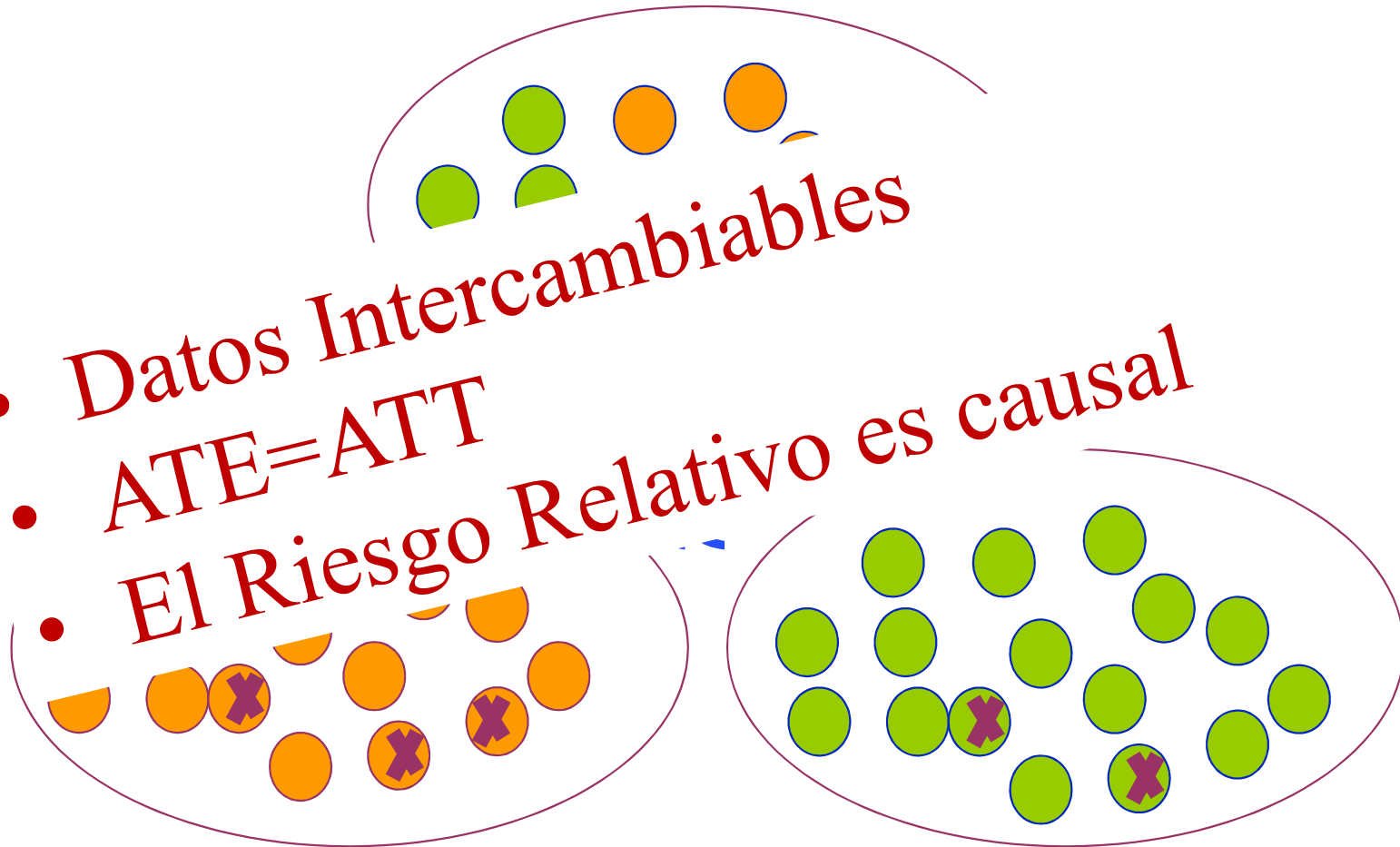
Nos gustaría saber que pasa si se intercambian los grupos

$$E(Y_{a=1}) = P(Y_{a=1} | A=0)P(A=0) + E(Y_{a=1} | A=1)P(A=1)$$

$$E(Y_{a=0}) = P(Y_{a=0} | A=0)P(A=0) + E(Y_{a=0} | A=1)P(A=1)$$

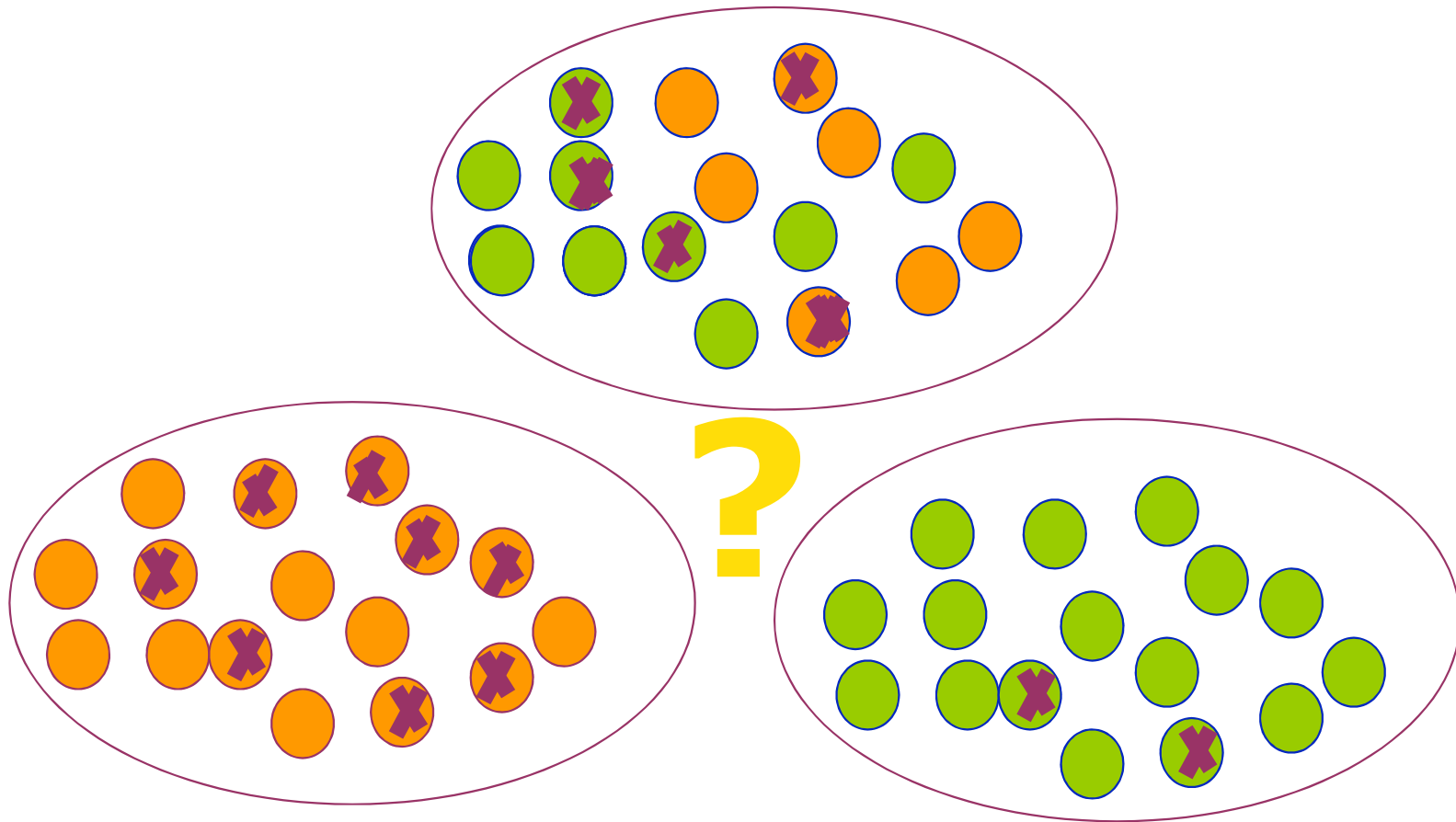
Ensayo Clínico

- Datos Intercambiables
- $ATE = ATT$
- El Riesgo Relativo es causal



Hiperpoblación contrafactual

Estudio Observacional



Hiperpoblación contrafactual

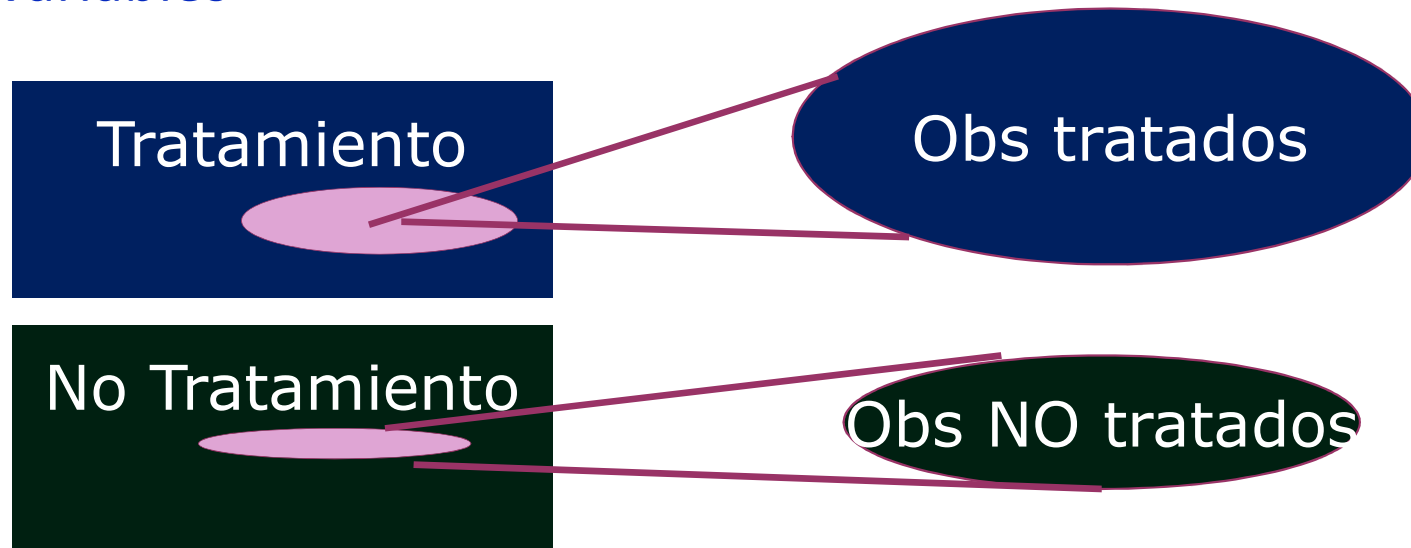
Estudio Observacional

- El tratamiento no se asigna aleatoriamente
- Los clínicos asignan el tratamiento en función del mejor pronóstico, cumplimiento, etc.
- La intercambiabilidad de las exposiciones no están garantizadas.
- El RR causal **no es igual** al RR observado
- El ATE **no es igual** al ATT

Solución intuitiva

Estudio observacional

- Los sujetos no son intercambiables dado que la probabilidad de estar tratado depende de las características de los sujetos.
- Serán intercambiables condicionados a tener las mismas variables



Reconstruir Hiperpoblación

Estudio Observacional

- Datos no tiene porque ser Intercambiables
- $ATE \neq ATT$
- El Riesgo Relativo NO es causal



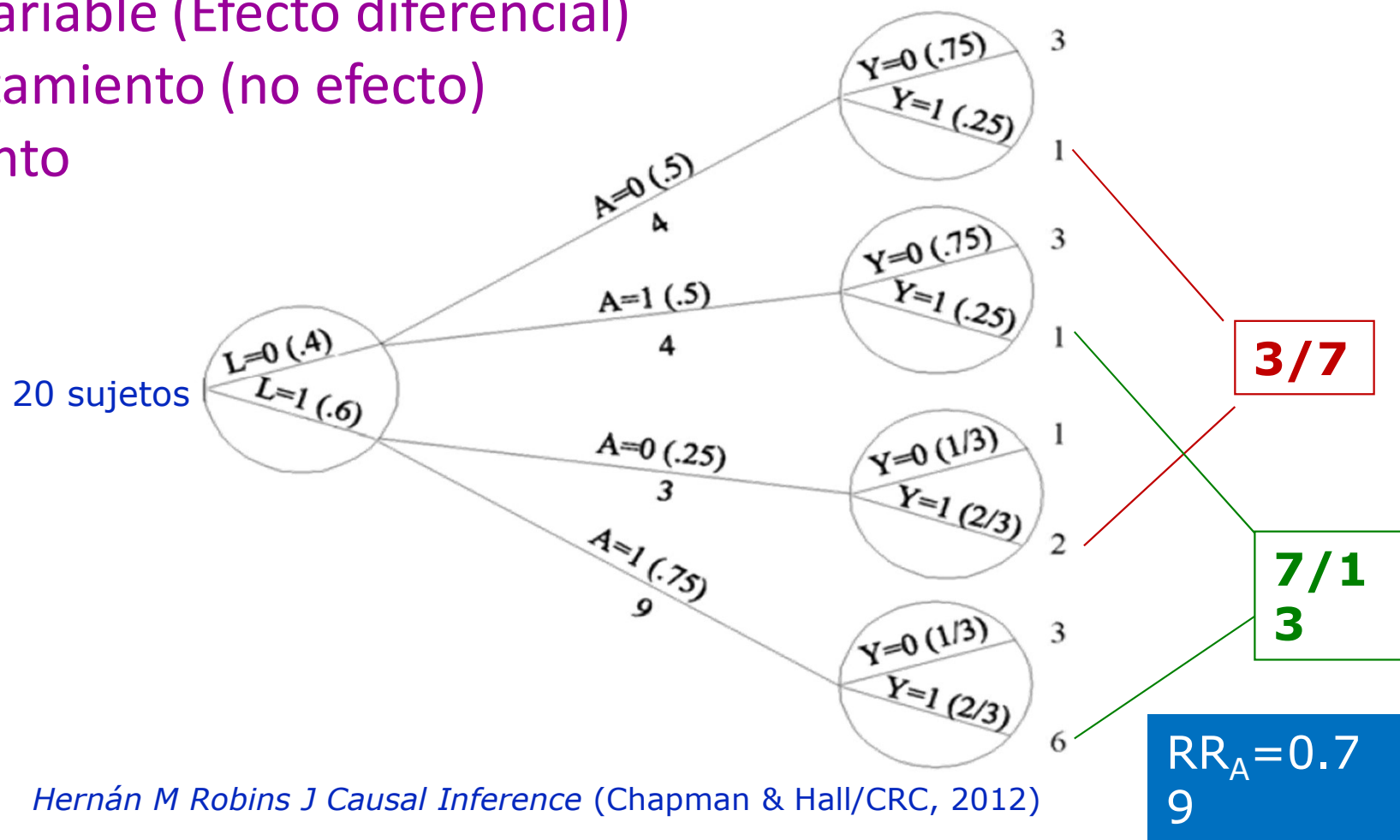
Hiperpoblación contrafactual

Estudio observacional

L= Covariable (Efecto diferencial)

A= Tratamiento (no efecto)

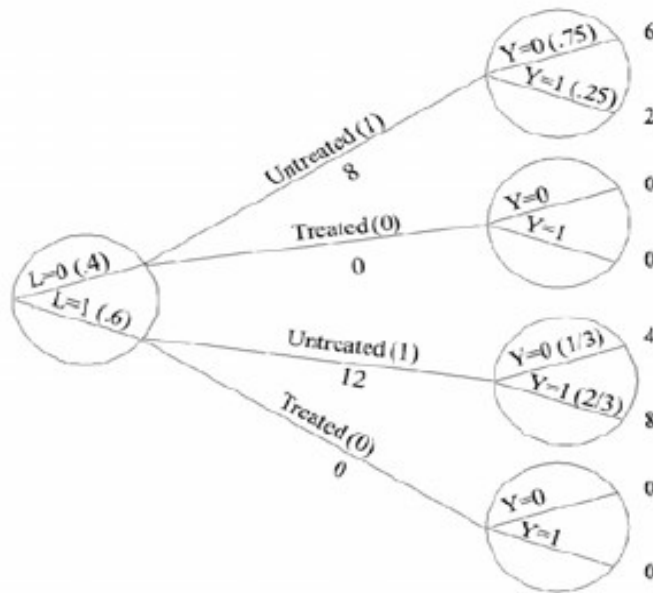
Y= Evento



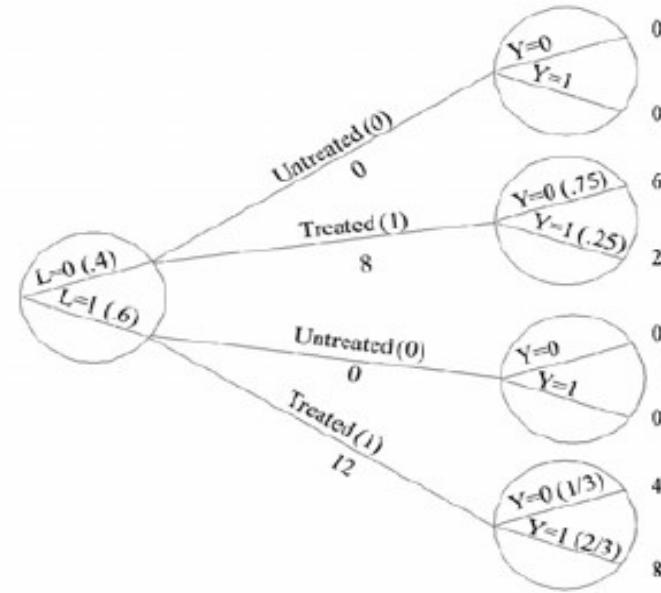
Hernán M Robins *J Causal Inference* (Chapman & Hall/CRC, 2012)

Riesgo Causal (Contrafactual)

Todos tratados



Todos NO tratados



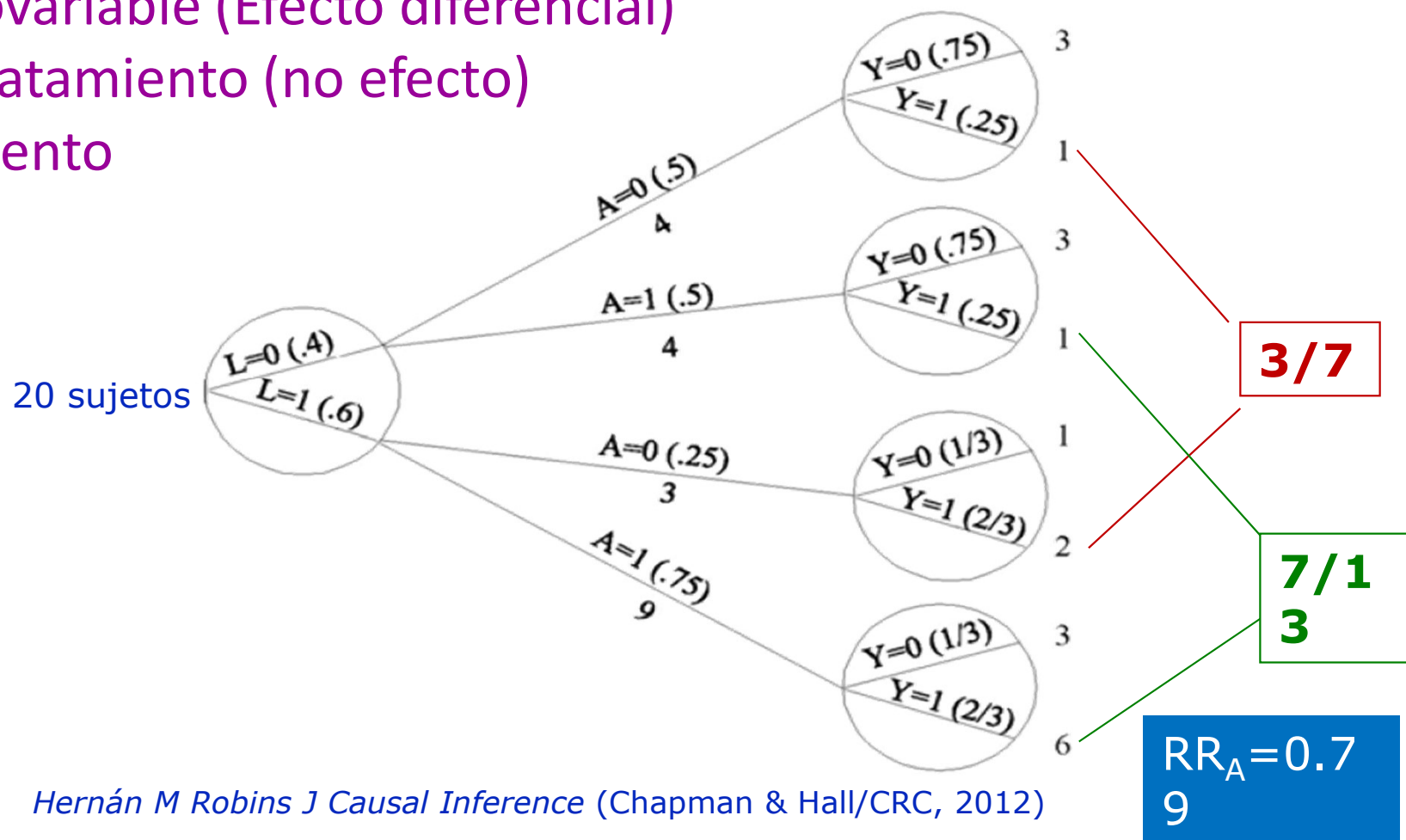
$$RR_c = (10/20) / (10/20) = 1$$

Estudio observacional

L= Covariable (Efecto diferencial)

A= Tratamiento (no efecto)

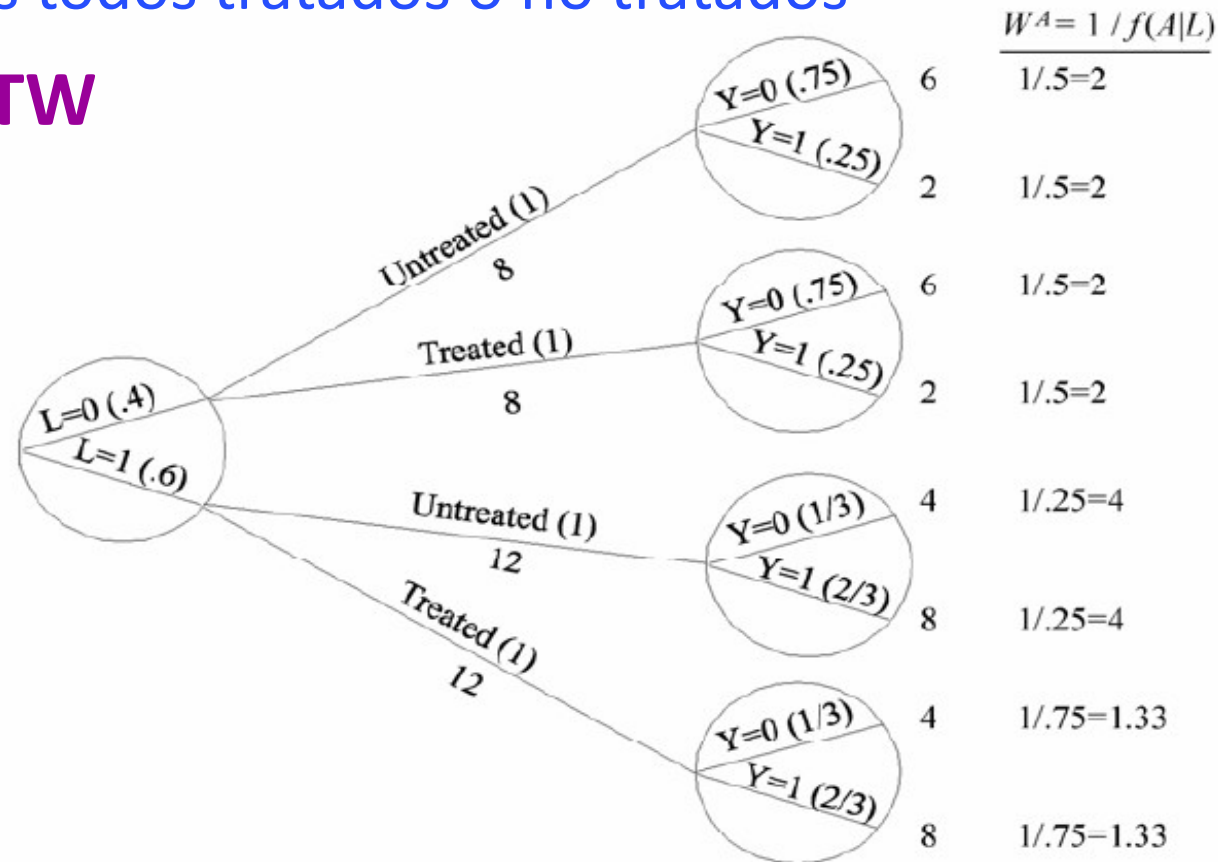
Y= Evento



Hernán M Robins J Causal Inference (Chapman & Hall/CRC, 2012)

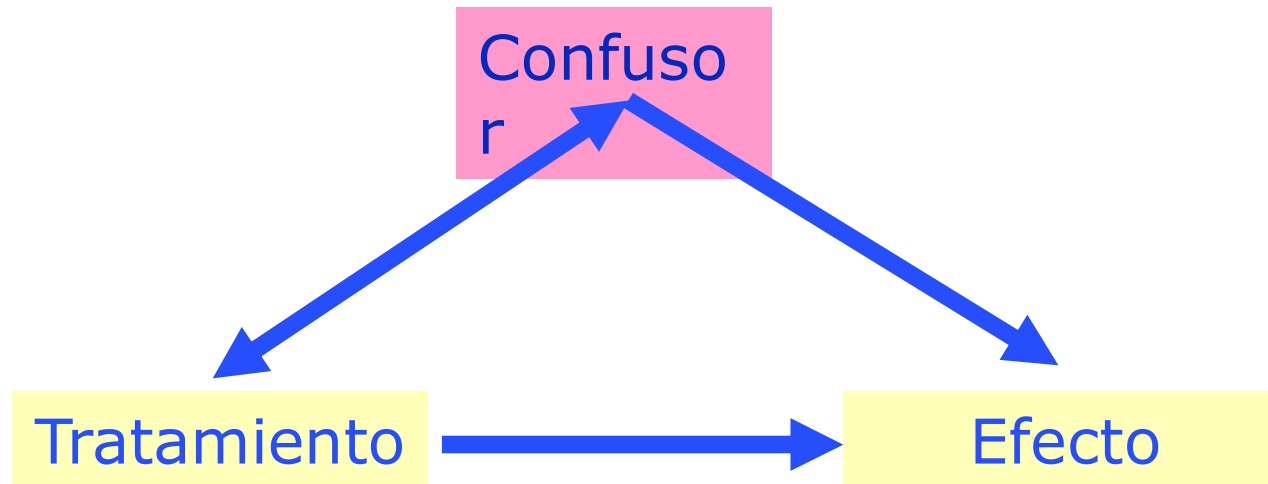
Pseudo población reconstruida con pesos Reconstruimos todos tratados o no tratados

IPTW



Hernán M Robins *J Causal Inference* (Chapman & Hall/CRC, 2012)

Confusión estándar



Control de la confusión

- **Método tradicional**

- Estratificación/Estandarización
- Modelos de regresión con covariables (**Reg. Multivariable**)

- **Alternativas**

- Estimar la probabilidad de estar tratado según las covariables consideradas (**Propensity Score**).
- Reanalizar **emparejando por PS, incluyendo PS** en el modelo, **estratificando por PS**
- Estimar la asociación entre el evento y el tratamiento utilizando pesos inversos a la probabilidad de tratar (**IPTW**)

Propensity Score

- PS=Propensity score = $P(A = 1 \mid L_0)$
- Probabilidad de estar asignado al tratamiento ó exposición según las características del individuo
- Se pueden calcular a partir de un modelo logístico para cada individuo

Propensity Score

Básicamente cuatro aproximaciones


- **Apareamiento** según nivel de PS por diferentes algoritmos (kernel,best neighbor,..)
- **Incluir PS** como covariable en el modelo
- **Estratificar** por valores del P(i.e quintiles) y efectuar un análisis estratificado
- Utilizar **pesos inversos** para recrear superpoblación

Cálculo de los pesos PS

- Modelo de **regression logística** multivariable con **respuesta el tratamiento/exposición**
- Se **incluyen variables basales** que se intentan equilibrar entre las relacionadas con el tratamiento y con la respuesta
- No importa el número de variables ya que no se pretende evaluar el efecto de estas variables sino **estimar la probabilidad de recibir el tratamiento**
- El mayor número de variables incluidas con interacciones, relaciones no lineales tienen mayor probabilidad de **reducir los sesgos de selección**
- El objetivo es **predecir para equilibrar no para discriminar**

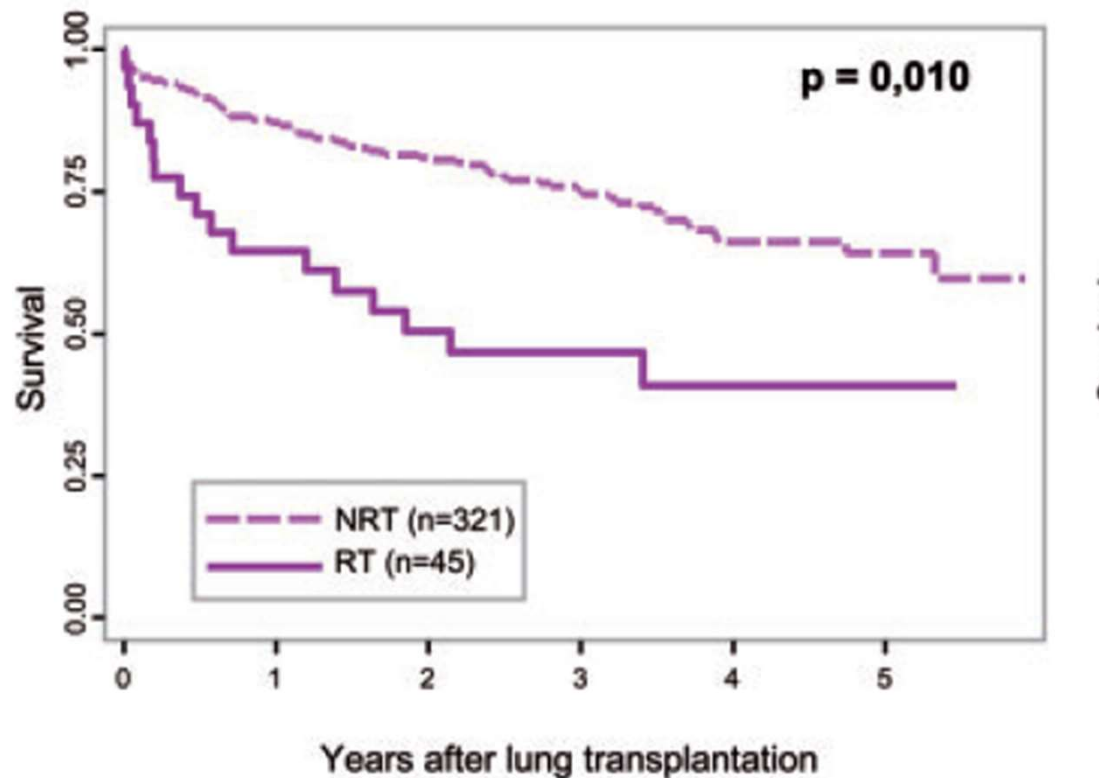
Ejemplo

Graft reduction surgery is associated with poorer outcome after lung transplantation: a single-centre propensity score-matched analysis

Pilar Montoya ^{a,*}, Irene Bello^a, Fernando Ascanio^a, Laura Romero^a, Javier Pérez^a, Joel Rosado^a, Leire Sánchez^a,

Judith Sacanell^b, Montserrat Ribas^c, Cristina Berastegui^d, Maria Deu ^a and Alberto Jáuregui^a

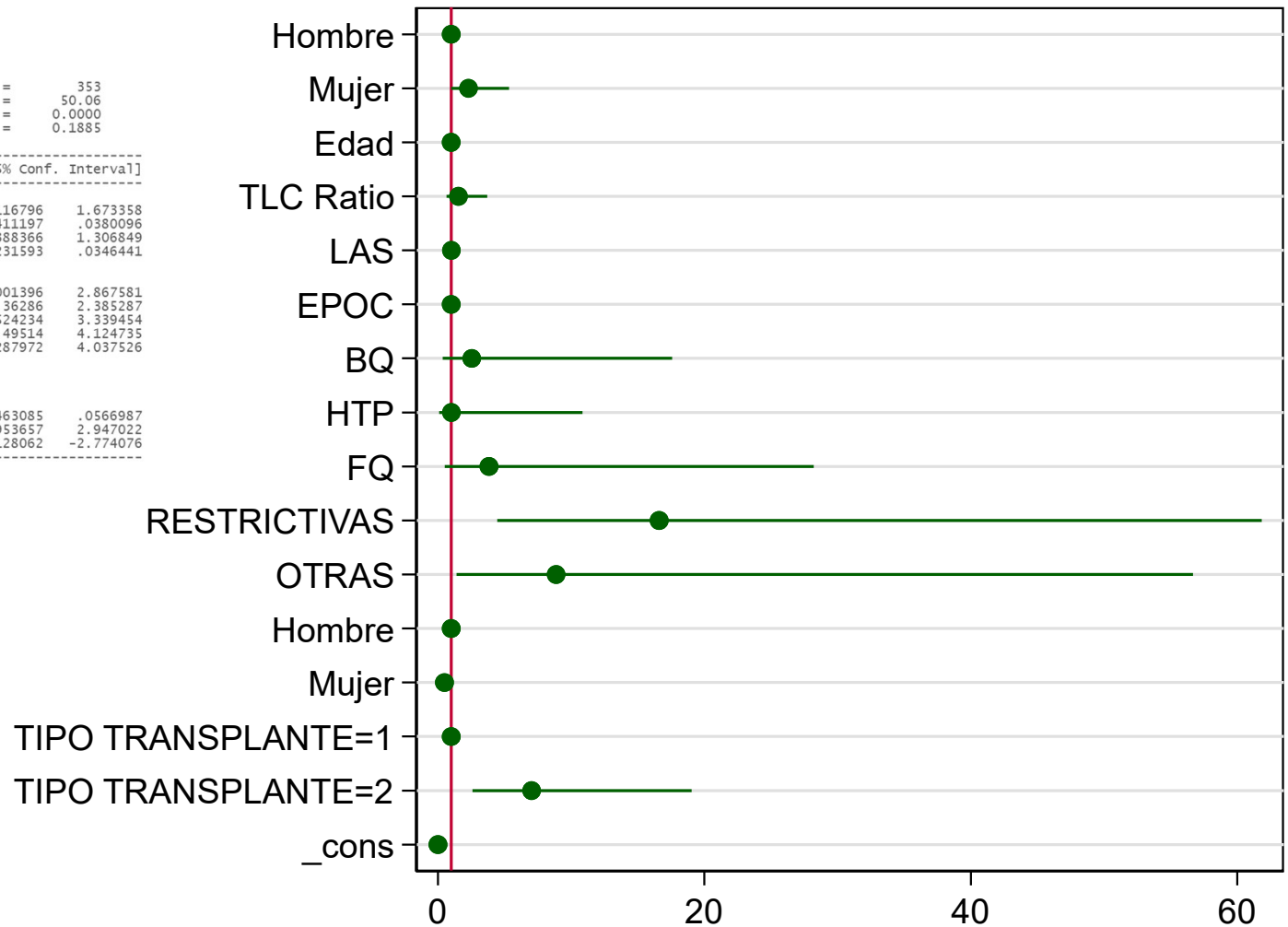
European Journal of Cardio-Thoracic Surgery 00 (2021) 1–8



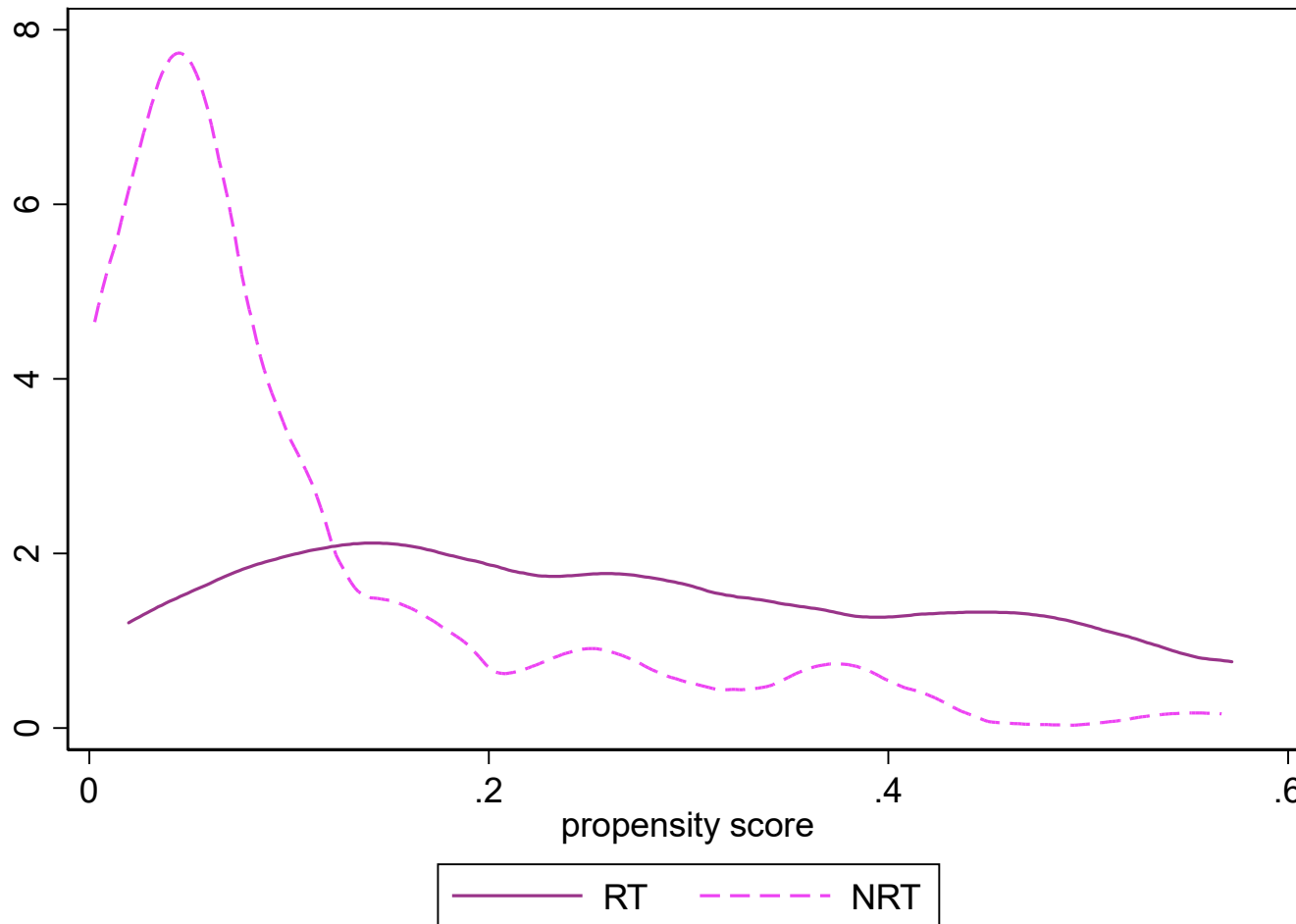
- 321 NRT = Non-reduced-size transplant patients
- 45 RT = Reduced-size transplant patients
- **HR=1.90 [1.15-3.13]**
- Características de pacientes diferentes según grupo de resección

Resultado modelo logístico

Logistic regression			Number of obs	=	353	
			LR chi2(11)	=	50.06	
			Prob > chi2	=	0.0000	
Log likelihood = -107.72796			Pseudo R2	=	0.1885	
<hr/>						
reseccion	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<hr/>						
sexo						
Mujer	.8308394	.4298645	1.93	0.053	-.0116796	1.673358
edad	-.001555	.0201864	-0.08	0.939	-.0411197	.0380096
tlcratio	.4340061	.4453361	0.97	0.330	-.4388366	1.306849
las	.0057424	.014746	0.39	0.697	-.0231593	.0346441
<hr/>						
patologia						
BQ	.9330927	.987002	0.95	0.344	-1.001396	2.867581
HTP	.0112137	1.211284	0.01	0.993	-2.36286	2.385287
FQ	1.343515	1.018355	1.32	0.187	-.6524234	3.339454
RESTRICTIVAS	2.809938	.6708274	4.19	0.000	1.49514	4.124735
OTRAS	2.183161	.9461216	2.31	0.021	.3287972	4.037526
LAM	0	(empty)				
Histiocitosis	0	(empty)				
<hr/>						
dsexo						
Mujer	-.7031931	.387707	-1.81	0.070	-1.463085	.0566987
2.tipotransplante	1.95034	.5085209	3.84	0.000	.953657	2.947022
_cons	-5.951069	1.620945	-3.67	0.000	-9.128062	-2.774076



Distribución de los propensity scores



Distribución de los propensity scores

Figure 1

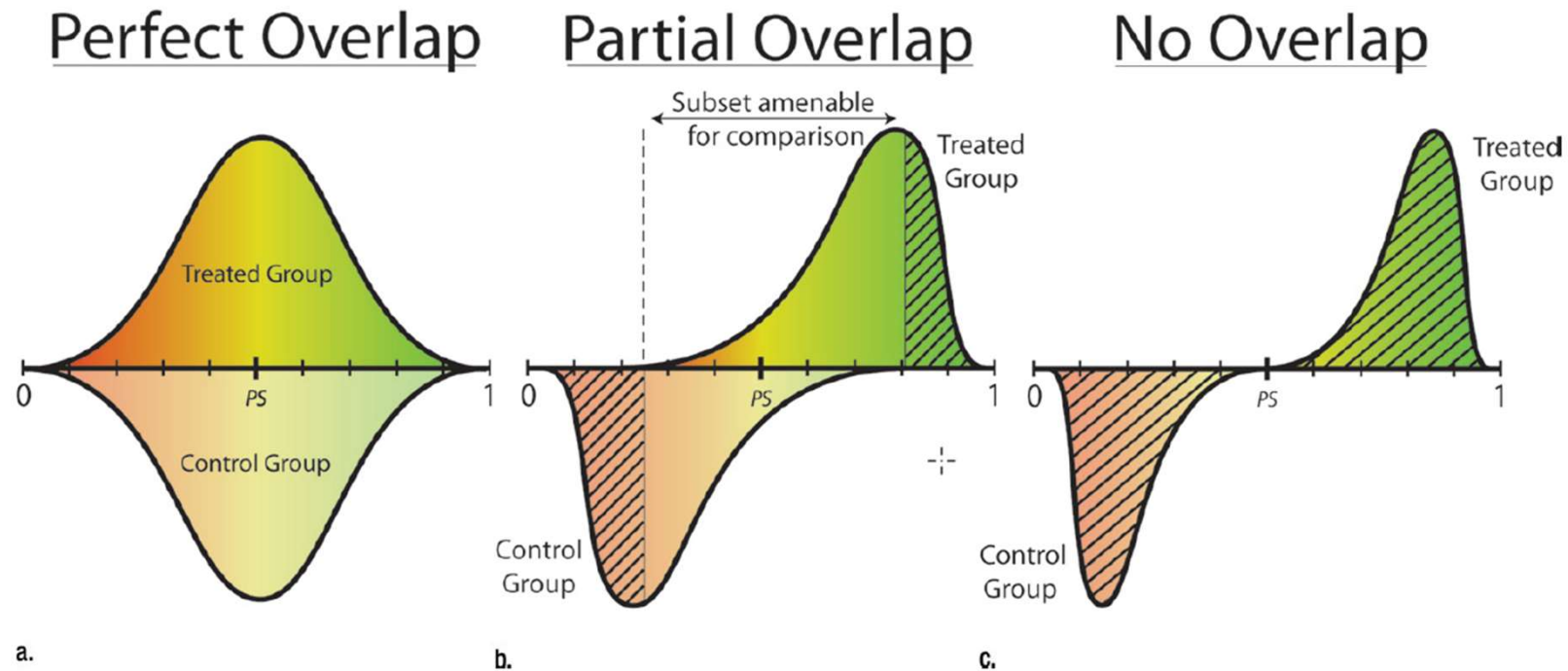


Figure 1: Hypothetical examples of treatment and control groups with (a) complete, (b) partial, and (c) no overlap in propensity score distribution. The x-axis shows the range of propensity scores (PS) from 0 (zero probability of a patient being given a specific treatment) to 1.0 (100% likelihood of a patient being given a specific treatment). Patients in the treatment group and those in the control group are above and below the x-axis, respectively. Shaded regions indicate patients in treatment and control groups who were not compared because their propensity scores did not overlap.

Behind The Numbers: Propensity Score Analysis McDonald et al*Radiology*: Volume 269: Number 3—December 2013

Figure 2

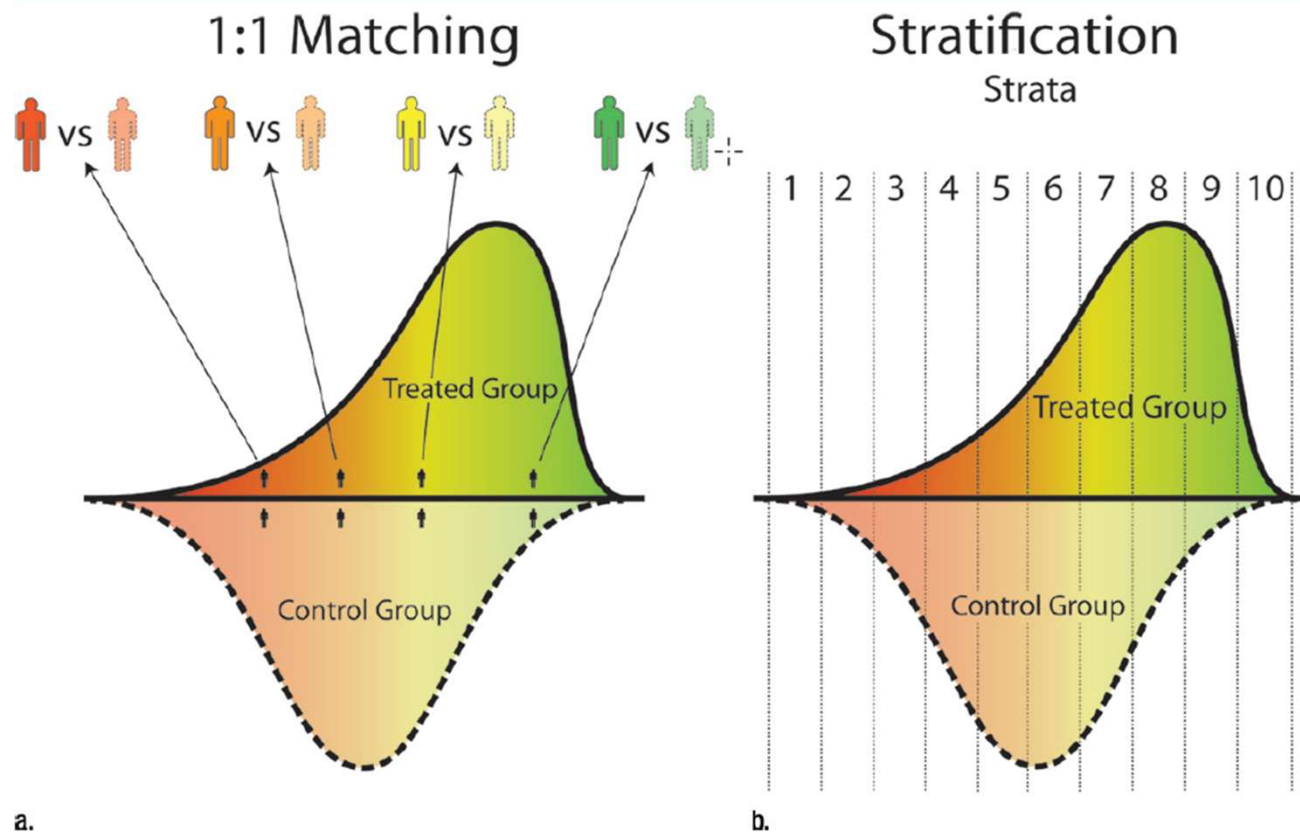
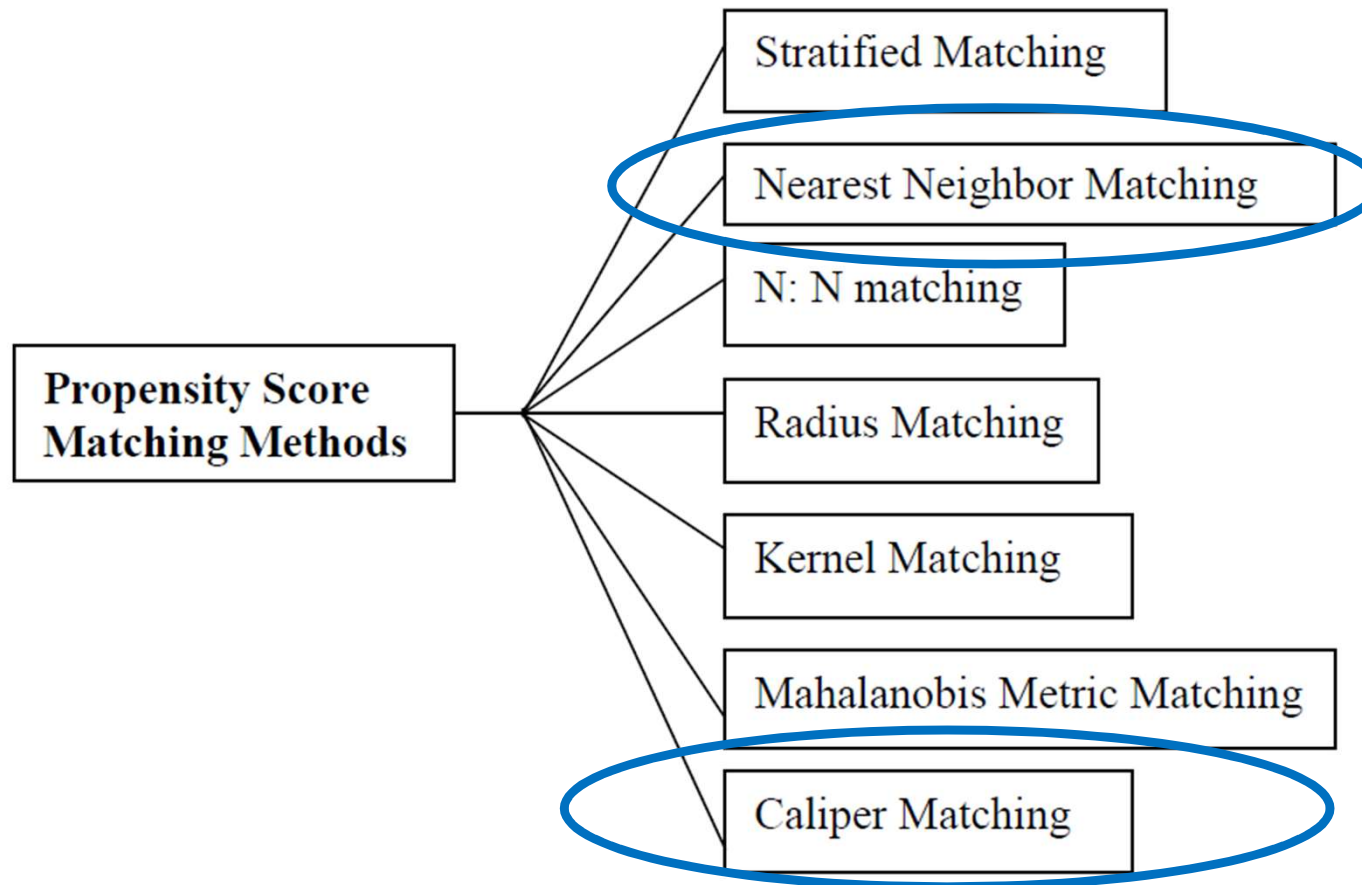


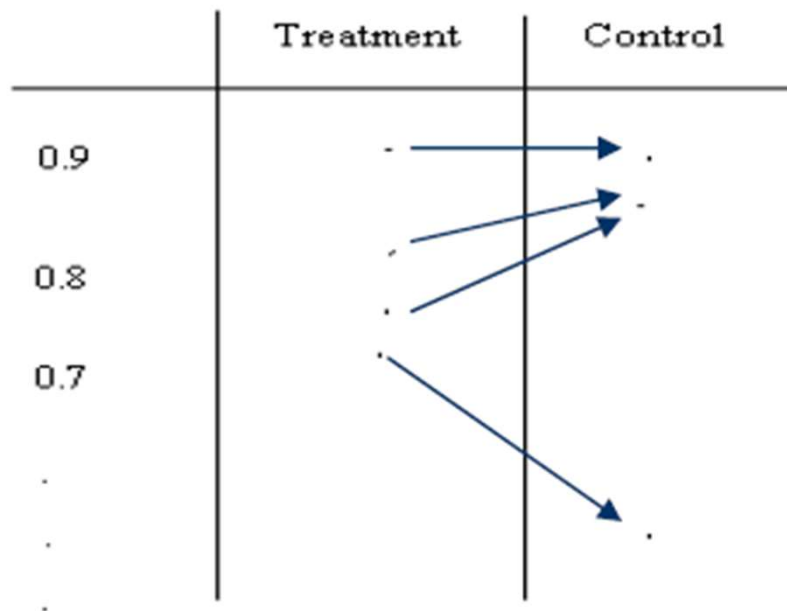
Figure 2: Methods of comparing patients in treatment and control groups after propensity score generation. **(a)** One patient in the treatment group is matched with one patient in the control group who has an identical or very similar propensity score. This 1:1 matching generates a cohort of an equal number of patients in treatment and control groups. **(b)** Patients in the treatment and control groups are stratified into deciles by propensity score. The resulting similar ranges of propensity scores are then compared between patients who underwent treatment and control subjects within each decile.

Behind The Numbers: Propensity Score Analysis McDonald et al*Radiology*: Volume 269: Number 3—December 2013

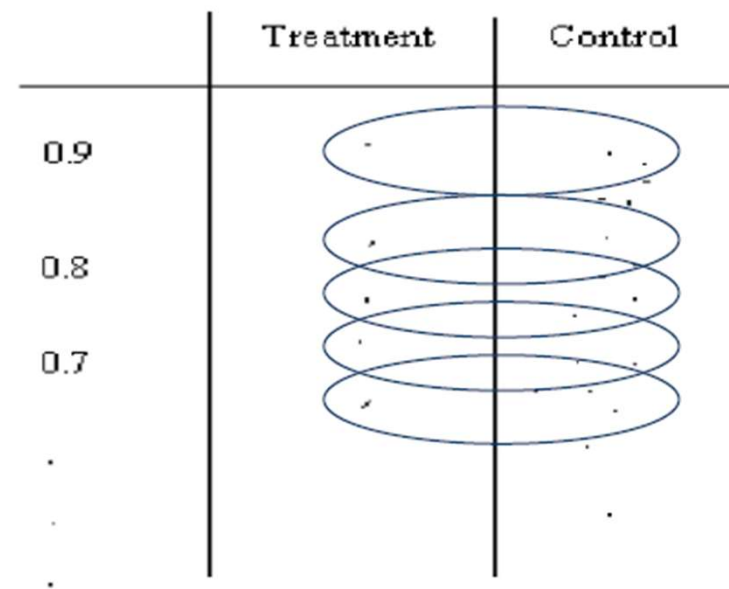
Propensity Matching



Tipo de apareamiento



Nearest Neighbor (NN)



Caliper & Radius

Diagnóstico de la calidad del apareamiento

- Hay que evaluar que las **variables de confusión** quedan **equilibradas** en los dos grupos apareados.
- No es suficiente con comparar las variables en los dos grupos y utilizar un p valor ya que al aparerar el tamaño muestral se ve reducido
- Hay que utilizar **medidas estandarizadas** como la diferencia media absoluta estandarizada

$$d = \left| \frac{(\bar{x}_t - \bar{x}_c)}{\sqrt{\frac{s_t^2 + s_c^2}{2}}} \right|$$

Medias

$$d = \left| \frac{(\hat{p}_t - \hat{p}_c)}{\sqrt{\frac{\hat{p}_t(1 - \hat{p}_t) + \hat{p}_c(1 - \hat{p}_c)}{2}}} \right|$$

Proporciones

Ejemplo

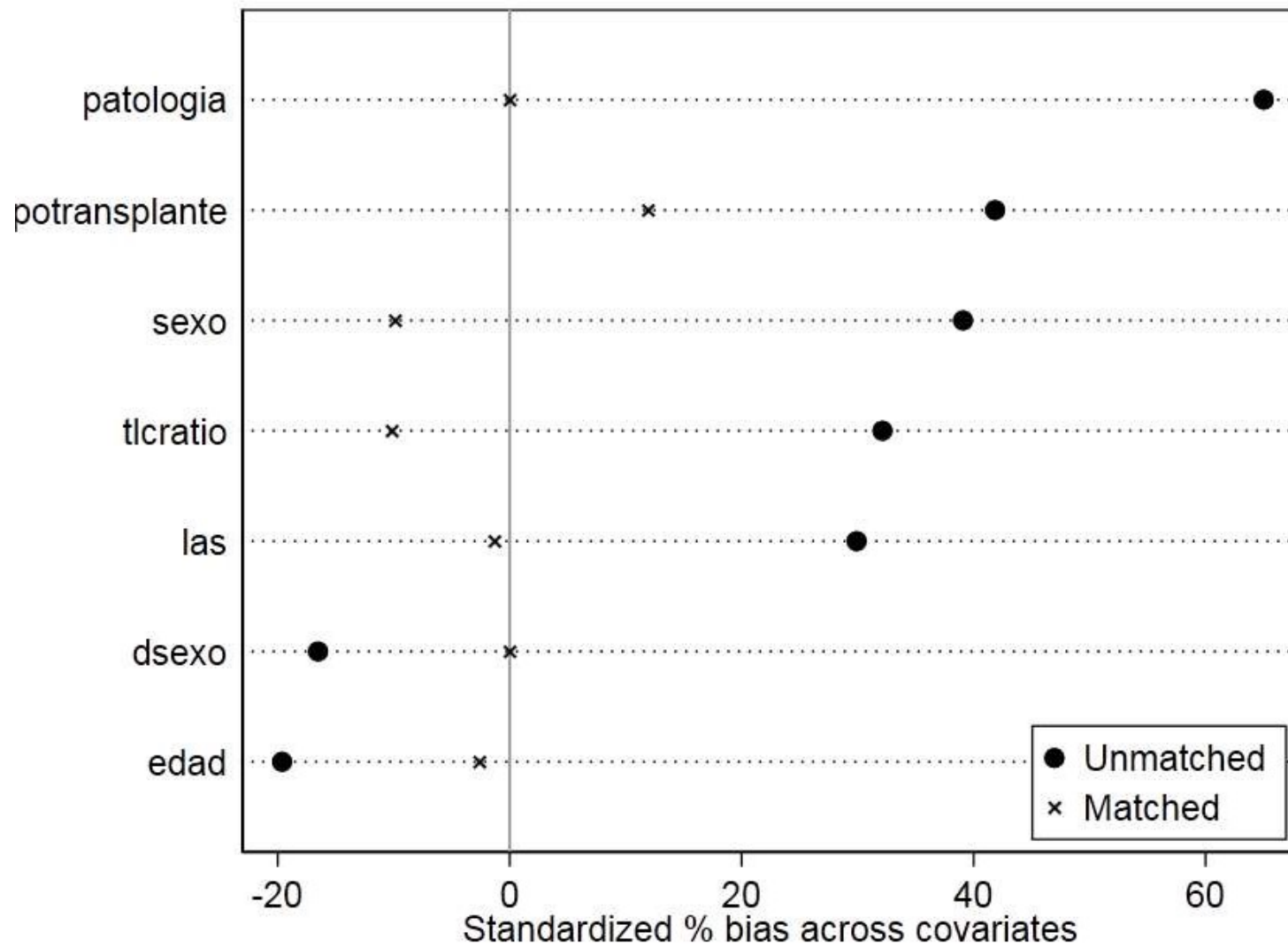
Variable	Unmatched Matched	Mean		%bias	%reduct bias	t-test		V(T)/ V(C)
		Treated	Control			t	p> t	
sexo	U	.61364	.42071	39.1		2.42	0.016	.
	M	.58537	.63415	-9.9	74.7	-0.45	0.656	.
edad	U	51.706	53.854	-19.6		-1.14	0.253	0.70
	M	51.698	51.982	-2.6	86.7	-0.11	0.916	0.54
tlcratio	U	.70455	.55016	32.1		1.94	0.053	.
	M	.68293	.73171	-10.2	68.4	-0.48	0.632	.
las	U	42.882	39.349	29.9		2.10	0.036	1.82
	M	42.407	42.561	-1.3	95.6	-0.05	0.959	1.10
patologia	U	4.5455	3.5307	65.0		3.58	0.000	0.47*
	M	4.5122	4.5122	0.0	100.0	0.00	1.000	1.03
dsexo	U	.43182	.51456	-16.5		-1.03	0.306	.
	M	.46341	.46341	0.0	100.0	0.00	1.000	.
tipotransplante	U	1.8636	1.6926	41.9		2.36	0.019	0.56
	M	1.8537	1.8049	11.9	71.5	0.58	0.563	0.80

* if variance ratio outside [0.55; 1.83] for U and [0.53; 1.88] for M

Sample	Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
Unmatched	0.167	44.31	0.000	34.9	32.1	114.5*	0.88	25
Matched	0.006	0.72	0.998	5.1	2.6	18.5	0.96	0

* if B>25%, R outside [0.5; 2]

Ejemplo



Analizar los datos apareados

- Para variable dicotómica hay que utilizar **regresión logística condicional**
- Para variables cuantitativas **test apareados** (t-test o Signo-rango Wilcoxon)
- Se pueden utilizar modelos **GEE** que tienen en cuenta la correlación entre medidas apareadas (varianza robusta)
- Para datos de supervivencia hay que utilizar modelo de **Cox con estimación robusta de la varianza** o un modelo de Cox estratificado en el apareamiento

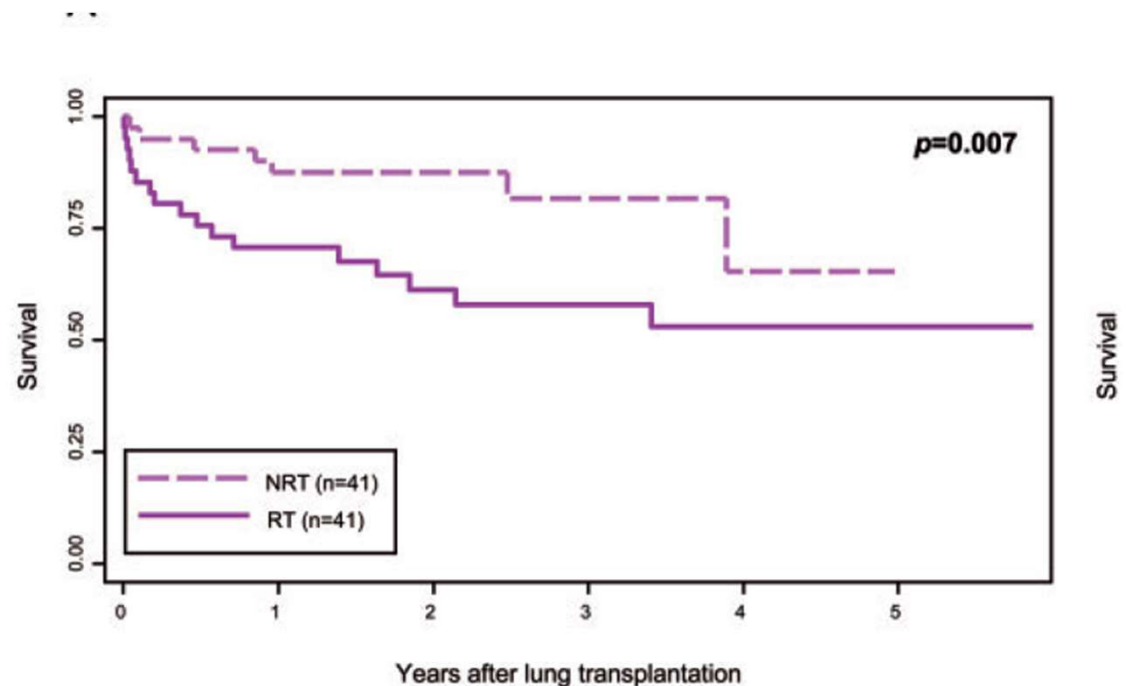
Ejemplo

- Se han emparejado pacientes con y sin resección que tenían la misma probabilidad de ser resecionados con un "caliper" de 0.02 Consiguiendo un total de 41 emparejamientos.
- Se calcula con los datos apareados el HR de supervivencia.

Crude **HR=1.90 [1.15-3.13]**

Table 4: Stratified univariable Cox regression for match-paired data

Variables	HR (95% CI)	P-value
Graft reduction	4.00 (1.34-11.96)	0.013
Anastomosis location	7.00 (1.50-30.00)	0.010



Inverse probability of treatment weighting (IPTW)

- Cuando la posibilidad de **emparejamiento** es complicada o muy limitada una opción es utilizar **IPTW**
- Se trata de **generar una población sintética** ponderando inversamente por la probabilidad de recibir el tratamiento
- De este modo las variables a nivel basal quedan equilibradas
- Los pesos son
 - $w_i = 1/P_{\text{score}_i}$ Si el sujeto i está tratado
 - $w_i = 1/(1-P_{\text{score}_i})$ Si el sujeto i NO está tratado
- Los pesos pueden ser inestables. En estos casos **hay que estabilizar los pesos**
- IPTW en propensity score es una aplicación de los **modelos marginales estructurales**

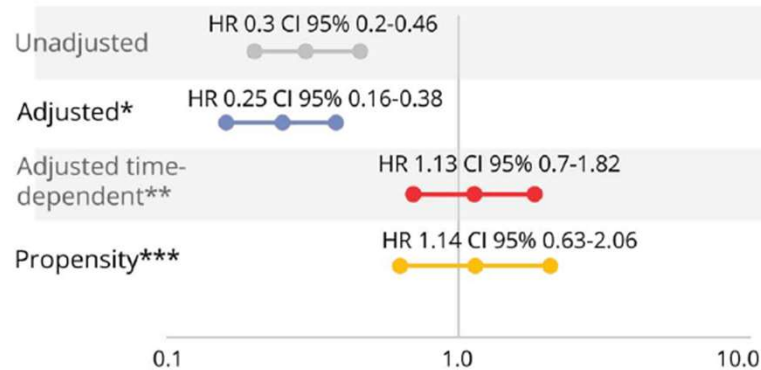
Menarche, pregnancies, and breastfeeding do not modify long-term prognosis in multiple sclerosis

María I. Zuluaga, MD, Susana Otero-Romero, MD, PhD, Alex Rovira, MD, Santiago Perez-Hoyos, BSc, PhD, Georgina Arrambide, MD, PhD, Laura Negrotto, MD, Ingrid Galán, MD, Jordi Río, MD, PhD, Manuel Comabella, MD, PhD, Carlos Nos, MD, María Jesús Arévalo, PhD, Angela Vidal-Jordana, MD, PhD, Joaquín Castilló, MD, PhD, Breogán Rodríguez, MD, Luciana Midaglia, MD, Patricia Mulero, MD, Raquel Mitjana, MD, Cristina Auger, MD, Jaume Sastre-Garriga, MD, PhD, Xavier Montalban, MD, PhD,* and Mar Tintoré, MD, PhD*

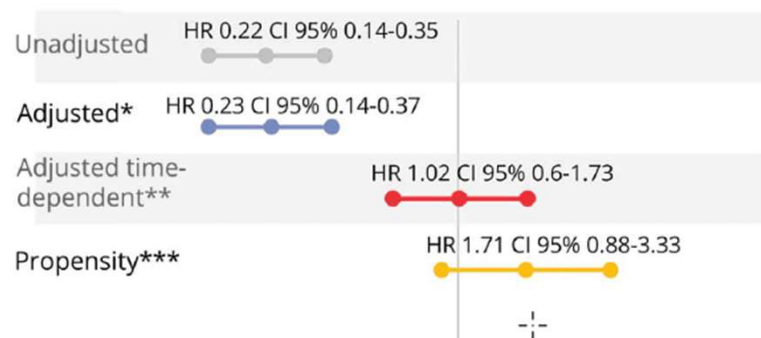
Correspondence
Dr. Tintoré
mtintore@cem-cat.org

Neurology® 2019;92:e1507-e1516. doi:10.1212/WNL.00000000000007178

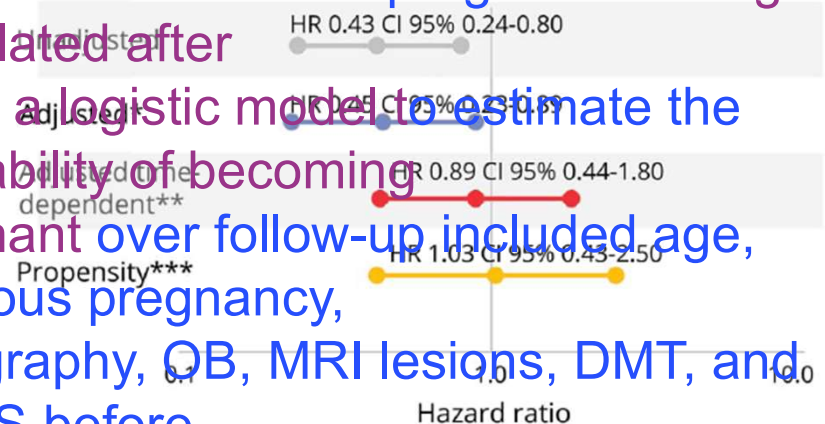
A. CDMS



B. McDonald 2010



C. EDSS 3-6



CI = confidence interval; HR = hazard ratio; OB = oligoclonal bands. *Adjusted by age at CIS, topography, OB, MRI, and disease-modifying treatment prior to a second attack (as a time-dependent variable). **Adjusted by age at CIS, topography, OB, MRI, and disease-modifying treatment prior to a second attack (as a time-dependent variable) and pregnancy as a time-dependent covariate. ***Propensity score model for pregnancy at any time using inverse probability weighting.

Ejemplo

by fitting a propensity score model for pregnancy at any using inverse probability weighting. These models controlled for selection bias because participants who became pregnant could exhibit different prognoses. Weights calculated after fitting a logistic model to estimate the probability of becoming pregnant over follow-up included age, previous pregnancy, topography, OB, MRI lesions, DMT, and CDMS before pregnancy.

5 Pasos para realizar un Análisis de PS

1. Identificar si un **Análisis de PS es viable y apropiado** para los datos disponibles
2. **Calcular los Propensity Scores** (Regresión logística outcome tratamiento y covariables confusoras basales)
3. **Seleccionar el método de PS** a utilizar (Matching, Estratificación, IPTW) y emparejar o calcular
4. **Asegurar el equilibrio** de las covariables tras ponderar o emparejar. Evaluar estabilidad de los pesos.
5. **Calcular los estimadores de efecto** en la muestra emparejada o ponderada siguiendo los métodos apropiados

Resumen

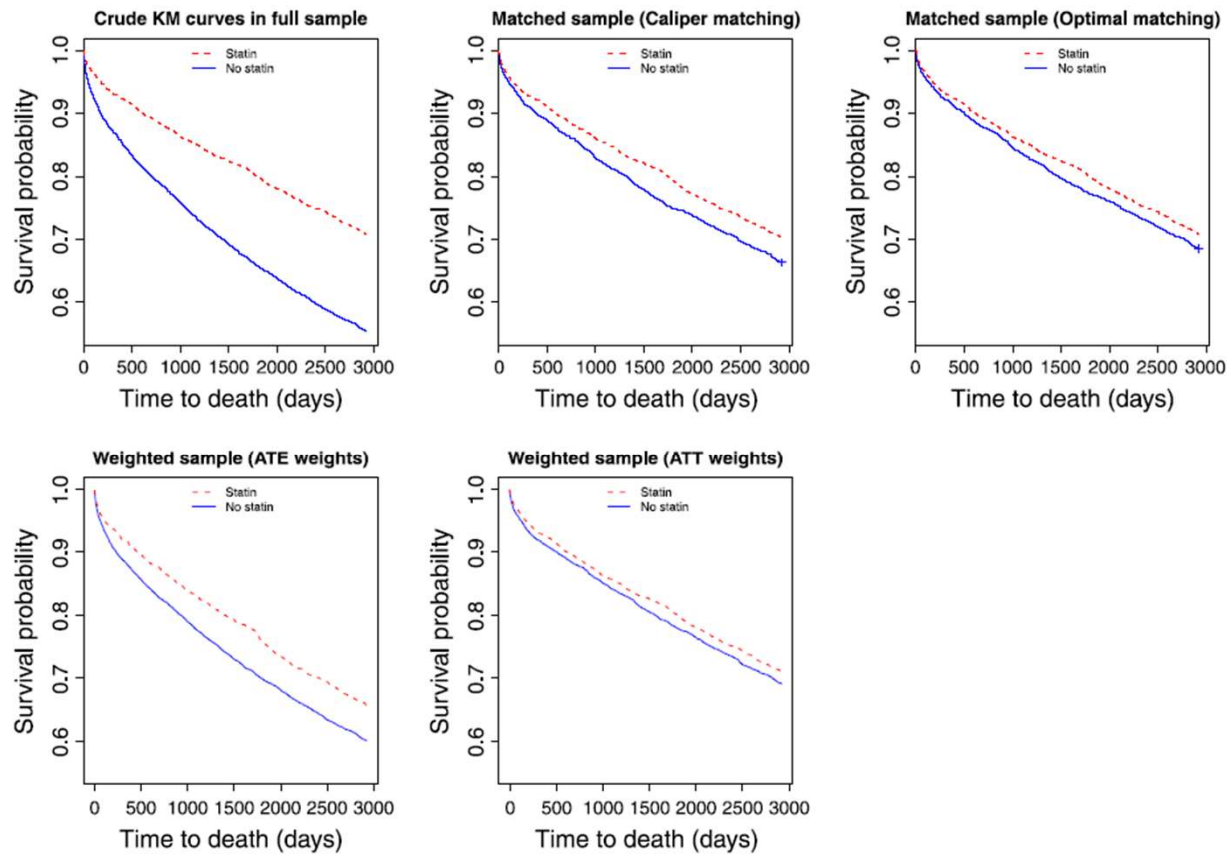


Figure 1. Kaplan-Meier survival curves obtained using different propensity score methods.

Software

Matching

Stata ps2match function

R Matchit Package

IPTW

Calculate weights and use

Weighted regression

Stata teffects ipw

R iptw

Recomendaciones

1

Table III. Recommendations for the use of propensity score methods with time-to-event outcomes.

Objectives	Propensity score matching (pair matching)	Inverse probability of treatment weighting using the propensity score
Estimand	ATT	ATE or ATT depending on the weights selected
Balance assessment	Compare distribution of baseline covariates between treated and untreated subjects in the matched sample.	Compare distribution of baseline covariates between treated and untreated subjects in the sample weighted by the inverse probability of treatment.
Estimate and report survival curves	Estimate Kaplan–Meier survival curves in treated and untreated subjects in the matched sample. Use stratified log-rank test to compare survival curves (stratify on matched sets).	Estimate adjusted Kaplan–Meier survival curves in the weighted sample. Use adjusted log-rank test to compare survival curves.
Estimate and report absolute reduction in the probability of an event occurring	From the estimated survival curves, estimate the absolute difference in the probability of an event occurring within a specified duration of follow-up.	From the estimated marginal survival curves, estimate the absolute difference in probability of an event occurring within a specified duration of follow-up.
Estimate relative change in the hazard of an event occurring	Use Cox proportional hazards model in the matched sample. Regress survival on an indicator variable for treatment selection. Use a robust variance estimator.	Use Cox proportional hazards model in the weighted sample. Regress survival on an indicator variable for treatment selection. Use a robust variance estimator.

ATT, average treatment effect for the treated; ATE, average treatment effect.

Bibliografia

- Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med*. 2014;33:1242–1258. DOI: [10.1002/sim.5984](https://doi.org/10.1002/sim.5984)
- Austin PC, Laupacis A. A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *International Journal of Biostatistics* 2011; **7**(1):Article 6. DOI:10.2202/1557-4679.1285
- Austin PC. A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behavioral Research* 2011; 46:1, 119-151 DOI: [10.1080/00273171.2011.540480](https://doi.org/10.1080/00273171.2011.540480)
- Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**(5):550–560. doi: 10.1097/00001648-200009000-00011.
- McDonald RJ, McDonald JS, Kallmes DF, Carter RE. Behind the Numbers: Propensity Score Analysis—A Primer for the Diagnostic Radiologist. *Radiology* 2013;269(3):640-45 DOI: [10.1148/radiol.13131465](https://doi.org/10.1148/radiol.13131465)
- Staffa SJ, Zurakowski D. Five Steps to Successfully Implement and Evaluate Propensity Score Matching in Clinical Research Studies. *Anesth Analg*. 2018 Oct;127(4):1066-1073. doi: 10.1213/ANE.0000000000002787.
- Zuluaga MI, Otero-Romero S, Rovira A, Perez-Hoyos S, Arrambide G, Negrotto L, Galán I, Río J, Comabella M, Nos C, Arévalo MJ, Vidal-Jordana A, Castilló J, Rodríguez B, Midaglia L, Mulero P, Mitjana R, Auger C, Sastre-Garriga J, Montalban X, Tintoré M. Menarche, pregnancies, and breastfeeding do not modify long-term prognosis in multiple sclerosis. *Neurology*. 2019 Mar 26;92(13):e1507-e1516. doi: 10.1212/WNL.0000000000007178.
- Montoya P, Bello I, Ascanio F, Romero L, Pérez J, Rosado J, Sánchez L, Sacanell J, Ribas M, Berastegui C, Deu M, Jáuregui A. Graft reduction surgery is associated with poorer outcome after lung transplantation: a single-centre propensity score-matched analysis. *Eur J Cardiothorac Surg*. 2021 Dec 1;60(6):1308-1315. doi: 10.1093/ejcts/ezab234..

Gràcies \ Gracias \ Thank you

