- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40

EDITORIALE

Il *propensity score*, uno strumento statistico sempre più usato in ambito cardiovascolare

Arnaldo Dimagli[1], Michele Di Mauro[2], Umberto Benedetto[1]

[1]*Bristol Heart Institute, University of Bristol, Bristol, UK*
[2]*Cardio-Thoracic Surgery Unit, Heart and Vascular Centre, Maastricht University Medical Centre (MUMC), Cardiovascular Research Institute Maastricht (CARIM), Maastricht, Olanda*

ADDENDA ONLINE

SCENARIO CLINICO

L'uso degli inibitori dell'enzima di conversione dell'angiotensina (ACE-inibitori) prima di un intervento di cardiochirurgia rimane controverso, con alcuni lavori che suggeriscono un aumentato rischio di insufficienza renale acuta[1]. Tuttavia, non esistono ancora studi randomizzati che hanno paragonato la strategia di somministrare ACE-inibitori fino al giorno dell'intervento con la strategia di sospenderli prima dell'intervento. Usando il database dell'Arterial Revascularization Trial (ART)[2], abbiamo analizzato l'incidenza di insufficienza renale acuta in 1948 pazienti che hanno ricevuto ACE-inibitori preoperativamente contro 977 che non hanno ricevuto ACE-inibitori.

Il *propensity score matching* richiede un algoritmo relativamente complesso e pertanto è necessario usare dei software statistici dedicati. Nel nostro caso utilizzeremo il programma statistico *R* e il pacchetto statistico *nonrandom.* Esistono anche altri pacchetti per eseguire il *propensity score matching* in R, come MatchIt e twang, mentre psmatch2 è disponibile in Stata e PS matching in SPSS.

1. Installazione e caricamento del pacchetto *nonrandom*

Dopo aver installato R, scaricabile attraverso l'URL nella bibliografia[3], è necessario installare e caricare il pacchetto *nonrandom*. Eseguire il download del pacchetto dal sito https://cran.r-project.org/src/contrib/Archive/nonrandom/ .

Dopo aver aperto R, dalla barra dei menu in alto selezionare "Tools", poi "Install packages…". Si aprirà una nuova finestra, CRAN Mirror, nell'opzione "Install from:" selezionare "Package archive file (.zip;.tar.gz)" e poi da "Browse" selezionare la cartella scaricata in formato .zip del file nonrandom. Cliccare su "Install". Il pacchetto verrà scaricato e installato. Dopodiché, per caricare il pacchetto bisognerà scrivere nella console il comando require(nonrandom) e così il pacchetto sarà caricato.

2. Preparazione e caricamento del dataset

Il dataset da utilizzare dovrà essere organizzato in modo che ogni riga corrisponda ad un caso e ogni colonna contenga una variabile.

Per caricare il dataset su R, si consiglia di utilizzare un file Excel salvato con estensione .csv. È importante che il file venga salvato nella stessa cartella dove è stata impostata la directory del progetto R in svolgimento. Il dataset da caricare non dovrà contenere dati mancanti. Per caricare il nostro file, chiamato "ART", occorre usare il seguente comando:

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40

```
mydata <- read.csv("ART.csv")
```

Questo comando legge, carica e rinomina il dataset "mydata" su R. Possiamo controllare il dataset caricato utilizzando il comando:

```
mydata[1:5,]
```

Questo ci permetterà di vedere il nome delle variabili e i primi cinque casi nel dataset (Figura S1). Il nostro dataset contiene 8 variabili, di cui "ACE" è la variabile discriminante che specifica se un determinato caso abbia ricevuto ACE-inibitori preoperativamente (ACE=1) o meno (ACE=0) e "AKI_postop" definisce il nostro outcome d'interesse (danno renale acuto postoperatorio). Le rimanenti sono l'età ("Age"), il sesso femminile ("FEMALE"), il valore di creatinina espresso in µmol/L ("Crea"), la presenza o assenza di diabete mellito ("Diabetes"), una frazione di eiezione ventricolare sinistra preoperatoria superiore o inferiore al 30% ("LVEF_pre"), e la presenza o assenza di fibrillazione atriale preoperatoria ("AF_pre").

3. Come calcolare il *propensity score*

Il primo comando da usare è pscore e permette di calcolare il PS sulla base di un modello di regressione logistica:

```
ps <- pscore(data = mydata, ACE ~ Age + FEMALE + Diabetes + Crea + LVEF_pre + AF_pre)
```

dove mydata indica il dataset, "ACE" è la variabile discriminante, e dopo la tilde ("~") sono elencate le variabili per le quali si vuole raggiungere un equilibrio tra i due gruppi. La scrittura "ps <-" permette di rinominare l'intero comando sotto la voce "ps" così che sarà più semplice un suo richiamo nei comandi successivi.
È poi importante andare a definire l'area di "supporto comune" (in inglese, *common support*), costruendo un grafico della distribuzione dei *propensity score* (Figura S1) con il comando:

```
plot.pscore(ps)
```

Per effettuare il match sulla base dei PS usiamo il comando ps.match:

```
match <- ps.match(ps, ratio = 1, caliper = 0.20, who.treated=0)
```

dove ps è l'oggetto derivato dal comando pscore precedente, ratio = 1 indica che ogni soggetto trattato sarà appaiato con un solo soggetto non trattato, caliper = 0.20 individua la massima distanza tra i PS di due soggetti che si vuole appaiare e se essa risulta maggiore del valore indicato tale appaiamento non potrà avvenire. Dato che nel nostro dataset, i soggetti nel gruppo sperimentale (ACE=1) erano in numero maggiore di quelli nel gruppo di controllo (ACE=0), è stato necessario specificare nel comando who.treated=0.
Una volta fatto ciò, il matching sarà stato eseguito e sarà possibile valutarne i risultati utilizzando il comando:

```
bal.table <- ps.balance(object = match, sel = c("Age", "LVEF_pre", "Crea", "AF_pre", "FEMALE", "Diabetes"),
method = "stand.diff", alpha = 10)
```

dove object = match rappresenta l'oggetto del nostro matching, sel indica le variabili considerate, method indica il metodo di confronto delle variabili tra i due gruppi e in questo caso è la standardized mean difference (SMD), alpha è il livello soglia percentuale rispetto al quale definire se l'equilibrio è raggiungo o meno, per

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40

cui se la SMD è inferiore al 10% la variabile sarà equilibrata tra i due gruppi.
Una volta eseguito il comando, richiamando bal.table si otterranno i dati dell'appaiamento. La Figura S2 mostra questi risultati ed in particolare si nota come la distribuzione delle variabili "Diabetes", "LVEF_pre", e "AF_pre" abbia raggiunto un equilibrio dopo il matching avendo ottenuto una SMD <10%. Questo è possibile vederlo anche con il *love plot* (Figura S3) utilizzando il comando:

```
plot.stdf(bal.table)
```

Il *love plot* mostra graficamente la variazione della SMD per ogni variabile prima e dopo il matching e nel presente studio mostra come per tutte le variabili la SMD sia inferiore al cut-off del 10%.

4. Ottenere il dataset dopo matching

Per ottenere il dataset contenente le coppie di pazienti appaiate sulla base del PS, si usa il seguente comando:

```
my.matched <- match$data.matched
```

All'interno di questo nuovo dataset, oltre alle variabili presenti già originariamente, vi sono due nuove colonne: "pscore" che contiene i PS calcolati per ogni paziente e "match.index" che identifica con lo stesso numero i pazienti di una coppia, permettendo di capire quale paziente del gruppo di controllo è stato appaiato con un determinato paziente del gruppo sperimentale.

5. Stima dell'effetto

Il campione dopo il matching va analizzato utilizzando strumenti statistici che rendano conto della natura appaiata dei dati. Per le variabili continue si può usare il t-test per campioni appaiati, mentre per le variabili binarie (sì/no, 0/1) si può adottare il test di McNemar. Alternativamente si può usare un modello lineare misto con la variabile trattamento (ACE-inibitori sì vs ACE-inibitori no) come effetto fisso e il numero di coppia (match.index) come effetto random.
Utilizzando il test di McNemar per confrontare l'incidenza della complicanza nel gruppo di controllo e nel gruppo sperimentale, si è ottenuta una differenza significativa con una maggiore incidenza di danno renale acuto nella popolazione esposta ad ACE-inibitori rispetto alla popolazione non esposta dopo matching (18.5% vs 14.0%, p<0.001). Per eseguire tale test, bisogna eseguire il comando:

```
mcnemar.test(my.matched$AKI_postop, my.matched$ACE)
```

in cui il primo argomento indica l'outcome di interesse e il secondo la variabile discriminante. Ogni argomento è preceduto dal nome del dataset a cui si fa riferimento (in questo caso il dataset dopo il matching, cioè my.matched), seguito dal simbolo "$", in quanto la funzione mcnemar.test non supporta l'argomento "data = " per cui occorre indicare per ogni variabile a quale dataset fa riferimento.
Il pacchetto lme4 permette di eseguire una regressione lineare mista:

```
glmer(AKI_postop ~ ACE + (1|match.index), data = my.matched, family = "binomial")
```

L'argomento family = "binomial" serve per definire un modello di regressione di tipo logistico, ovvero quando l'outcome di riferimento è binario (danno renale acuto postoperatorio presente o assente). Nel nostro studio è risultato un OR 1.33 (IC 95% 1.03-1.72), per cui l'assunzione di ACE-inibitori preoperatoria è associata allo sviluppo di danno renale acuto postoperatorio.

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40

BIBLIOGRAFIA

1. Yacoub R, Patel N, Lohr JW, Rajagopalan S, et al. Acute kidney injury and death associated with renin angiotensin system blockade in cardiothoracic surgery: a meta-analysis of observational studies. Am J Kidney Dis 2013;62:1077-86.

2. Taggart DP, Lees B, Gray A, Altman DG, Flather M, Channon K; ART Investigators. Protocol for the Arterial Revascularisation Trial (ART). A randomised trial to compare survival following bilateral versus single internal mammary grafting in coronary revascularisation [ISRCTN46552265]. Trials. 2006;7:7

3. R Core Team (2014). R: A language and environment for statistical computing. (3.0.3) [Computer software]. Vienna, Austria: Foundation for Statistical Computing. http://www.R-project.org/.

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40

Figura S1. Risultato del codice mydata[1:5,] che rappresenta i primi 5 casi del dataset con tutte le variabili presenti.

```
> mydata[1:5,]
    Age FEMALE Crea Diabetes LVEF_pre AF_pre ACE AKI_postop
1 66.66      0  122        0        0      0   1          0
2 63.41      1   80        0        0      0   1          0
3 64.84      0  100        0        0      0   0          0
4 66.37      0  104        1        0      0   1          0
5 50.74      0  110        0        0      0   1          0
```

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40

Figura S2. Controllo del bilanciamento tra le variabili dopo il matching, utilizzando il comando bal.table. La differenza tra i due gruppi è calcolata come standadized mean difference e i due gruppi sono equilibrati per tale variabile se la standardized mean difference è inferiore al 10%.

```
> bal.table

 Summary of balance check:

                       before: no balance (0) before: balance (1)
after: no balance (0)                       0                   0
after: balance (1)                          3                   3


 Covariates not completely tested: ---


 Detailed balance check (overall):

             Age Crea Diabetes FEMALE LVEF_pre AF_pre
table.before   1    1        0      1        0      0
table.after    1    1        1      1        1      1


 Detailed balance check:
 [Standardized differences (cut point: 10)]

            Age  Crea Diabetes FEMALE LVEF_pre AF_pre
Before    1.209 5.636   21.716  1.152   22.646 13.762
------    ----- -----    -----  -----    -----  -----
After     1.288  0.54     2.17  2.394    1.334  1.371

---------  ----  ----     ----   ----     ----   ----
Scale       num   num      bin    bin      bin    bin
```

- Copyright - Il Pensiero Scientifico Editore downloaded by IP 216.73.216.151 Mon, 14 Jul 2025, 21:46:40
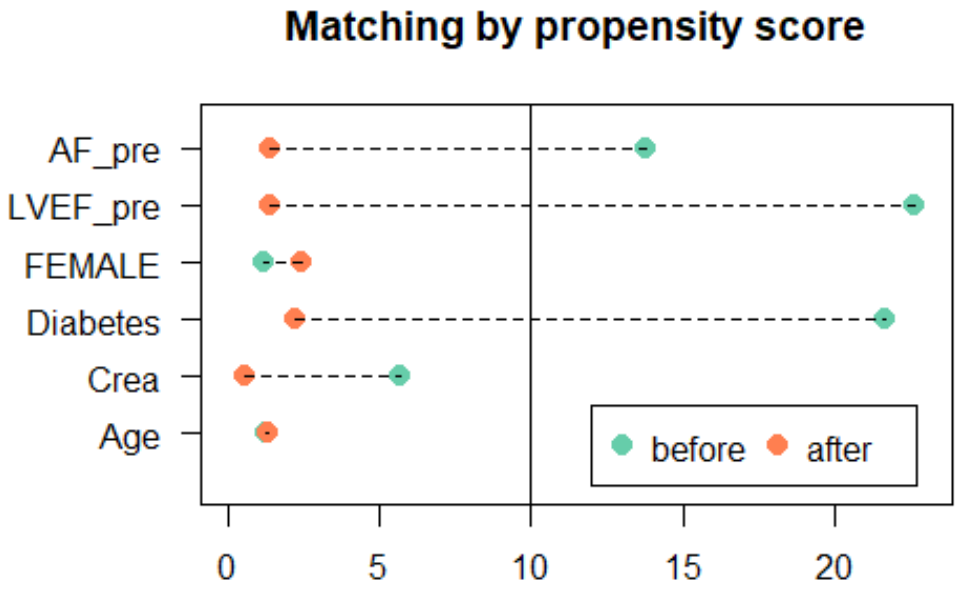
Figura S3. Love plot che mostra la variazione della standardized mean difference prima (in blu) e dopo (in rosso) il matching.