

傾向分數方法於藥物流行病學研究之應用

慈濟大學公衛系 溫淑惠

根據 Cochran & Chambers (1965)文章提出觀察性研究(observational study)為無法透過隨機指派病人接受治療的方式，評估某一暴露因素跟結果變項間的因果關係(cause-and-effect relationship)之研究；在 Guo & Fraser (2015)書中也提到觀察性研究有兩個特點，(一)觀察性研究關心治療的效果(treatment effect)，與(二)觀察性研究的資料不是經由隨機實驗而來。目前觀察性研究常見資料來源有健康調查資料、醫院病歷或健保資料庫等。以下為簡化說明，我以吃藥與否兩組人的治療效果為例，觀察性研究目的為了解病人吃藥的效果，但無法透過隨機化(randomization)指派誰該吃藥；缺點則為吃藥與不吃藥兩組人的期初特性(baseline characteristics)可能異質性很高而造成兩人治療效果無法公平比較。尤其是藥物流行病學議題，病人是否該吃藥常會受到病人本身的狀況、疾病嚴重度、醫師開藥習慣甚至是疾病診斷時的時間(像是藥物使用指引的改變或是藥物下市停用)等因素影響，亦即 confounding by indication 的問題，因此必須透過適當且嚴謹的觀察性研究，方可適切評估藥物的治療效果或其副作用。

為降低觀察性研究的偏差(bias)，必須透過對干擾因素做適當的配對(match)、分層(stratification)或調整(adjustment)來處理干擾效應(confounding effect)；在藥物流行病學領域，可能的干擾因素通常很多(諸如年齡、性別、共病等)，因此透過傾向分數(propensity score, 簡稱PS)法進行上述控制干擾的方法就很常見。PS 最早由 Rosenbaum & Rubin 於 1983 年提出，主要想法是將多個干擾因素(期初收集的變數)轉換成一個機率值，此機率值即稱為 PS，可白話解釋為“傾向接受治療”的可能性，PS 數值越大顯示病人吃藥的可能性越高。為方便說明，先定義符號如下：主要感興趣的結果變項(primary outcome variable) Y 為一兩分類的變數(binary variable)可分為有($Y = 1$)、無($Y = 0$)發生事件，以 X 代表暴露變項也是兩分類($X = 1$ 為吃藥組, $X = 0$ 為未吃藥組), Z_1, \dots, Z_k 為 k 個潛在的干擾因

素；PS 可利用邏輯斯迴歸(logistic regression)

$$\text{logit}(P(X = 1 | Z_1, \dots, Z_k)) = \gamma_0 + \sum_{i=1}^k \gamma_i Z_i \text{ 估計}$$

而得 PS 為 $[1 + \exp(-(\hat{\gamma}_0 + \sum_{i=1}^k \hat{\gamma}_i Z_i))]^{-1}$ ，其中 $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_k$ 為迴歸係數估計值。現行諸多統計軟體皆可以邏輯斯迴歸功能快速進行 PS 之估計，像是 SAS 軟體的 proc logistic 語法，Stata 軟體的 logistic 語法，R 軟體的 glm 語法中設定 family = binomial。若欲進一步比較吃藥與否兩組人的 PS 分佈，Stata 軟體的 pscore (需寫語法”ssc install pscore”自行安裝此功能)可進行 PS 分組後吃藥與否兩組人的 Z_1, \dots, Z_k 是否相近，或是 R 軟體的 PSAgraphics(需自行下載此套件)可提供兩組人的 PS 視覺化比較。

近十年許多觀察性研究使用 PS 進行干擾因素的控制有快速增加的趨勢 (Gayat et al. 2010)，PS 可透過配對、分層、調整及加權(weighting)等四種方法來處理干擾效應(Faries et al. 2010)，簡述如下：

- (1) 若使用 PS 配對(以下簡稱為 PS-M 法)時，是依據吃藥組的 PS，決定適當的間隔比如 PS 加減 0.01 或是 $\text{logit}(\text{PS})$ 的 0.2 倍標準差的範圍內，在未吃藥組尋找具相近的 PS 為其配對；缺點為治療與否兩組的 PS 分布在頭尾兩端通常較無重疊，不易找到合適的配對者，因此會喪失部份比例的樣本，若未能成功配對的樣本過多也會導致推論對象的特性改變；然而 PS-M 法得以推論的對象會限縮到吃藥組的特性，也就是只能估計吃藥者的治療效果(average treatment effect for the treated, 簡稱 ATT)。以 Seeger et al. (2005)的文章分析使用降血脂藥物與心肌梗塞的關係為例，利用期初 52 個干擾因素建構 PS 後用於配對(未能成功配對的樣本近三成)，PS-M 法發現降血脂藥物會將低心肌梗塞發生風險比(hazard ratio, $\text{HR}=0.69$, $95\% \text{ CI}=0.52-0.93$)，與未進行 PS 配對的分析結果($\text{HR}=2.09$, $95\% \text{ CI}=1.58-2.76$)有相反的結果，顯見用藥兩組不公平的比較會導致錯誤的結論。現行統計軟體也提供 PS-M 法功能，像是 Stata

軟體的 psmatch2，R 軟體的 MatchIt 套件，讀者若有興趣可自行查找相關語法。

(2) PS 也可進行分層分析(以下簡稱為 PS-S 法)，比如利用 PS 的 20、40、60、80 百分位數將樣本分成 5 層，針對各層的吃藥與否兩組進行分析，爾後再針對各層的分析結果採取加權的做法得到整體吃藥效果的估計；雖然 PS-S 法可使用全部的樣本，但在 PS 頭尾兩端較易產生吃藥與否兩組樣本數不平衡的狀況，仍有可能造成吃藥效果估計不佳。

(3) 利用 PS 調整的作法則是採取迴歸分析的策略(以下簡稱為 PS-R 法)，將結果變項 Y 視為反應變數，解釋變數通常僅需放入 X 與 PS 值 (原始 PS 值或是以 PS 百分位數分組)，調整 PS 後估計 X 與 Y 的相關性，PS-R 法模式可寫成 $\log it(P(Y | X, PS)) = \alpha + \beta_1 X + \beta_2 PS$ 。在 Stürmer et al. (2006) 文章回顧 2003 年以前使用 PS 方法的文獻，發現 PS-R 法與傳統迴歸模式

$(\log it(P(Y | X, Z)) = \alpha + \beta_1 X + \sum_{i=1}^k \beta_{i+1} Z_i)$ 兩種方法的吃藥效果估計結果近九成是一致的；研究也顯示不一致的部分以 PS-R 法的估計結果比較保守(Faries et al. 2010)。然而，PS-R 法的優點是此模式須納入的解釋變數個數比傳統迴歸模式少很多，較可能符合迴歸分析的假設(assumptions)，而且 PS 估計較不易受到 model misspecification 的影響(Faries et al. 2010)。

(4) 除上述三種作法，PS 也可以加權方式(以下簡稱為 PS-W 法)進行干擾因素的控制。有兩種較常用的加權方式，第一種稱作 weighted by odds 或稱為 standardized mortality ratio (SMR) weighting，這類的方法是將吃藥組權重設為 1，未吃藥組的權重則為 $PS/(1-PS)$ ，這樣的設定方式可以估計吃藥者的治療效果(即 ATT)；第二種稱作 inverse probability of treatment weighting(簡稱 IPTW)，吃藥組權重設為 $1/PS$ ，未吃藥組的權重則為 $1/(1-PS)$ ，這樣的設定方式可以估計一般族群平均治療效果(average treatment effect, 簡稱 ATE)。

Y 與 X 的相關即可透過上述兩種 PS-W 的加權方式，以加權的邏輯斯迴歸分析估計吃藥的效果。

Ali et al. (2015) 以 PubMed 查詢 2011/12-2012/5 期間共計 296 篇醫學文獻使用 PS 方法，以 PS-M 法最常被使用(68.9%)，其次依序為 PS-R 法(20.9%)，PS-S 法(13.9%)及 PS-W 法(IPTW, 7.1%)。系統性回顧文獻也指出四種不同 PS 法應用存在差異且估計結果也有不同；Dahabreh et al.(2012)利用 Medline 查詢 2011/2 以前使用 PS 的觀察性研究進行急性冠狀動脈症候群(acute coronary syndrome) 的治療效果的文獻並由作者篩選隨機控制試驗(randomized controlled trial, 簡稱 RCT)以比較兩者估計結果者共計 21 篇，其中有 18 篇使用 PS-R 法的方法，使用 PS 法估計的結果會比 RCT 較易得到治療有效的結果，也就是較會高估治療效果。而 Zhang et al. (2014) 以 PubMed 查詢 2013/4 前重症醫學議題，共計 20 篇使用 PS 的文獻其中 14 篇使用 PS-M 法，同樣與適合的 RCT 結果比較，發現 PS 估計的結果反而比 RCT 較不易得到治療有效的結果，也就是較會低估治療效果。

實務上如何考量研究欲估計的效應值(effect size)以及資料的特性採取較合適的分析策略也是很重要的。學者 Peter Austin 利用模擬研究評估上述四種 PS 的表現，模擬常見治療(接受治療機率接近 0.5)及非罕見結果變項(發生事件機率接近 0.25)的資料，分別討論在估計效應值為風險差異值(risk difference, 簡稱 RD)、相對風險(relative risk, 簡稱 RR) 以及勝算比(odds ratio, 簡稱 OR)的表現 (Austin 2007; Austin et al. 2007; Austin 2008; Austin 2010)，其中 OR 包含 Marginal OR 及 Conditional OR 兩種，Marginal OR 為母體平均(population-average)效應，亦即吃藥發生結果事件之 OR；Conditional OR 為個體層次(subject-specific)的效應，亦即給定其他解釋變數下，吃藥發生結果事件之 OR (Austin 2007)。文獻中評估估計效果的指標包含偏差(bias)、均方誤(mean square error, MSE)、95% CI 覆蓋率(coverage probability of 95% CI)等；Peter Austin 模擬常見結果變項的狀況

下，四種不同 PS 法隨著估計的效應值不同，估計的偏差也有差異；像是估計 marginal OR 時以 PS-M 法為首選，但在估計 conditional OR 時，反而傳統迴歸方法表現最好；RD 的估計以 PS-W 表現最佳。

不過 PS-M 法的應用若是針對罕見暴露藥物流病議題時，可能會喪失太多非暴露組的個案進行疾病風險估計。Desai et al. (2017) 採用實例分析及模擬研究的方式，提出罕見暴露(0.5%-1%)時若能結合 PS-S 法與加權迴歸估計法可讓吃藥與否兩組人的期初干擾因素相近而且會有最佳的疾病風險估計。主要想法是先排除吃藥(暴露)與否兩組人 PS 分佈未重疊處，再根據吃藥組的 PS 進行 PS-S 法分成 50 組，爾後根據未吃藥組的人之 PS 分數再分到上述 50 個吃藥組的分組裡。接著對每一細分組(已包含吃藥者及未吃藥者)則採取加權迴歸模式，權重設定吃藥者為 1，未吃藥者為 $=(\text{該組吃藥者占總吃藥者比例})/(\text{該組未吃藥者占總未吃藥者比例})$ ，此概念也就是 SMR weighting，換句話說可以估計 ATT。作者也提供其 SAS 語法在此網站 <http://www.drugapi.org/dope-downloads/>，讀者若有興趣可自行查找相關語法。本文僅簡略介紹傾向分數方法於藥物流行病學研究之四種常見之應用方式：PS-M 法、PS-R 法、PS-S 法及 PS-W，研究者應依據研究議題之屬性選擇合適之分析方法，以期善用資料之訊息進行正確之疾病風險估計。

References

1. Austin, PC (2007). The performance of different propensity score methods for estimating marginal odds ratios. *Statistics in medicine*, 26(16), 3078-3094.
2. Austin, PC, Grootendorst, P, Normand, SL, Anderson, GM (2007). Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Statistics in medicine*, 26(4), 754-768.
3. Austin, PC (2008). The performance of different propensity-score methods for estimating relative risks. *Journal of clinical epidemiology*, 61(6), 537-545.
4. Austin PC (2010). The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Statistics in medicine*, 29(20), 2137-2148.

5. Cochran WG, Chambers SP (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society. Series A (General)*, 128, 234-266
6. Rosenbaum PR, Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
7. Seeger JD, Williams PL, Walker AM (2005) An application of propensity score matching using claims data. *Pharmacoepidemiology and drug safety*, 14: 465–476
8. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S (2006) A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of clinical epidemiology*, 59(5), 437-e1.
9. Gayat E, Pirracchio R, Resche-Rigon M, Mebazaa A, Mary JY, Porcher R (2010). Propensity scores in intensive care and anaesthesiology literature: a systematic review. *Intensive Care Med*;36(12): 1993-2003.
10. Faries DE, Obenchain R, Haro JM, Leon, AC (2010) Analysis of observational health care data using SAS. SAS Institute.
11. Seeger JD, Williams PL, Walker AM (2005) An application of propensity score matching using claims data. *Pharmacoepidemiology and drug safety*, 14: 465–476
12. Ali MS, Groenwold RH, Belitser SV, et al. (2015). Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of clinical epidemiology*, 68(2), 112-121.
13. Dahabreh IJ, Sheldrick RC, Paulus JK, et al. (2012). Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. *European heart journal*, 33(15), 1893-1901.
14. Zhang Z, Ni H, Xu X (2014). Observational studies using propensity score analysis underestimated the effect sizes in critical care medicine. *Journal of clinical epidemiology*, 67(8), 932-939.
15. Ross ME, Kreider AR, Huang YS, Matone M, Rubin DM, Localio AR (2014) Propensity score methods for analyzing observational data like randomized experiments: challenges and solutions for rare outcomes and exposures. *American journal of Epidemiology*, 181, 989-995.
16. Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF (2017). A propensity-score-based fine stratification approach for confounding adjustment when exposure is infrequent. *Epidemiology*, 28(2), 249-257.