การวิเคราะห์คะแนนความโน้มเอียงในการวิจัยทางวิทยาศาสตร์สุขภาพ

สิริมา มงคลสัมฤทธิ์ วท.ด (ระบาดวิทยาคลินิก)[1]*

นนท์ธิยา หอมขำ ปร.ด (สถิติ)[1]

[1] คณะสาธารณสุขศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ศูนย์รังสิต ปทุมธานี ประเทศไทย

* ผู้ติดต่อ, อีเมล์ tu.sirima@gmail.com

Vajira Med J. 2018; 62(1): 33-42

http://dx.doi.org/10.14456/vmj.2018.4

บทคัดย่อ

การวิจัยทางด้านวิทยาศาสตร์สุขภาพส่วนใหญ่เป็นรูปแบบการศึกษาเชิงสังเกต และปัญหาสำคัญที่พบ คือ เรื่องการจัดการตัวแปรกวนและตัวแปรร่วมจำนวนมาก ซึ่งอาจมีผลกระทบต่อความน่าเชื่อถือของผลการศึกษา เมื่อนำตัวแปรจำนวนมากเหล่านั้นเข้าโมเดลวิเคราะห์ข้อมูลทางสถิติในคราวเดียวกัน ปัจจุบันวิธีการคำนวณคะแนนความโน้มเอียงเป็นวิธีการทางสถิติที่มีการกล่าวขานและใช้กันเพิ่มขึ้นในรูปแบบการศึกษาเชิงสังเกต คะแนนความโน้มเอียงใช้หลักการคำนวณคะแนนความน่าจะเป็นโดยอาศัยข้อมูลของตัวแปรร่วมหลายตัวให้เหลือเพียง 1 ตัวแปร เพื่อบอกโอกาสในการได้รับสิ่งแทรกแซงของกลุ่มตัวอย่างแต่ละคน และนำคะแนนความโน้มเอียงที่คำนวณได้ไปใช้ร่วมกับวิธีการอื่นๆ เช่น การจับคู่ การแบ่งกลุ่มเป็นชั้น และการวิเคราะห์ถดถอย ทั้งนี้คะแนนความโน้มเอียงจะไม่มีประโยชน์สำหรับการวิจัยที่ออกแบบการศึกษาไม่ดี หรือไม่มีการเก็บตัวแปรกวนที่สำคัญไว้ในการศึกษา แต่ในทางตรงกันข้ามกลับเป็นเครื่องมือที่ดีในกรณีที่การศึกษามีการออกแบบการศึกษามาเป็นอย่างดี

คำสำคัญ คะแนนความโน้มเอียง ตัวแปรกวน การศึกษาเชิงสังเกต อคติ

Propensity Score Analysis in Health Science Research

Sirima Mongkolsomlit PhD (Clinical Epidemiology)[1*]
Nontiya Homkham PhD (Statistics)[1]

[1] Faculty of Public health, Thammasat University, Rangsit campus, Pathumthani, Thailand
* Corresponding author, e-mail address: tu.sirima@gmail.com
Vajira Med J. 2018;62(1): 33-42
http://dx.doi.org/10.14456/vmj.2018.4

Abstract

Research in health sciences mainly takes the form of observational studies. The major problem encountered is the management of confounding variables and many covariate variables. This affects the reliability of the data analysis on a number of variables in the model at the same time. Currently, propensity score is a statistical method that is well-known and used increasingly in the design of observational studies. Propensity score has been used to summarize data of several variables as one covariate; the propensity score estimates the probability of covariates based on each individual of received intervention. Furthermore, the propensity score has been used with other analysis techniques such as matching, stratification and regression analysis. The propensity score is not appropriate for poor research or potentially confounding variables. On the other hand, the propensity score is a good tool for good research management.

Key words: propensity score, confounding, observational study, bias

บทนำ

เป็นที่ทราบกันว่ารูปแบบการศึกษาแบบ randomized controlled trials (RCTs) ได้รับการยอมรับว่าสามารถควบคุมการเกิดอคติ (bias) ได้ดีที่สุดซึ่งนับว่าเป็น “gold standard” ในการออกแบบการศึกษาเชิงทดลอง เมื่อผู้วิจัยต้องการศึกษาเพื่อประเมินผลของการรักษาหรือค้นหาปัจจัยต่าง ๆ หรือพิสูจน์สิ่งทดลองว่ามีผลอย่างไร เนื่องจากมีขั้นตอนการสุ่มกลุ่มตัวอย่างที่เป็นระบบ (random allocation) ทำให้โอกาสของการเกิดตัวแปรกวนหรือตัวแปรร่วมอื่น ๆ มีการกระจายเท่าๆ กัน หรือใกล้เคียงทั้งในกลุ่มศึกษาและกลุ่มควบคุม แต่ก็ยังคงมีบางตัวแปรที่มีโอกาสที่อาจจะเกิดขึ้นไม่เท่ากัน ถึงแม้จะมีการสุ่มกลุ่มตัวอย่างที่เป็นระบบ (random allocation) ซึ่งเป็นลักษณะที่เกิดขึ้นโดยธรรมชาติ[1]

การวิจัยทางด้านวิทยาศาสตร์สุขภาพส่วนใหญ่ที่ใช้รูปแบบการศึกษากึ่งทดลอง (quasi experimental) และรูปแบบการศึกษาเชิงสังเกต (observational study) เนื่องจากมีข้อจำกัดทางด้านจริยธรรมวิจัยที่ไม่สามารถทำการศึกษาแบบทดลองได้ หรือทรัพยากรไม่เอื้อต่อการทำการวิจัยแบบทดลอง[2,3] ซึ่งรูปแบบการศึกษาดังกล่าวมักมีผลกระทบเนื่องจากตัวแปรกวนที่ทำให้เกิดอคติขึ้นเนื่องมาจากอิทธิพลของลักษณะพื้นฐานของกลุ่มตัวอย่างหรือวิธีการเลือกกลุ่มตัวอย่าง ซึ่งหากเปรียบเทียบลักษณะทั่วไปของกลุ่มศึกษาและกลุ่มควบคุมในรูปแบบการศึกษาเชิงสังเกตจะพบว่ามีลักษณะพื้นฐานระหว่างกลุ่มตัวอย่างที่แตกต่างกัน ซึ่งแหล่งที่ทำให้เกิดอคติในรูปแบบการศึกษาเชิงสังเกตมีหลายแหล่ง[4,5] ได้แก่ 1) ตัวแปรกวน (confounder) ทำให้เกิดความไม่น่าเชื่อถือของผลการศึกษา ประกอบด้วยสองรูปแบบ คือ confounding by indication และ confounding by contraindication 2) อคติที่เกิดขึ้นเนื่องจากการจดจำข้อมูลในอดีตไม่ได้ (recall bias) หรือ 3) วิธีการวัดที่แตกต่างกัน (detection bias) ซึ่งในงานวิจัยที่เป็นรูปแบบเชิงสังเกตมักพบว่ามีอคติที่เรียกว่า confounding by indicator เช่น ในผู้ป่วยเบาหวานที่เป็นความดันโลหิตสูง ตรวจพบโปรตีนในปัสสาวะ มีแนวโน้มที่จะได้รับยากลุ่ม RAS blocker มากกว่าคนที่ไม่เป็นความดันโลหิตสูงหรือตรวจไม่พบโปรตีนในปัสสาวะ ถ้าหากผู้วิจัยต้องการศึกษาการเสียชีวิตจากโรคไตวายในผู้ป่วยเบาหวานระหว่างกลุ่มที่ได้รับและไม่ได้รับ ยากลุ่ม RAS blocker โดยใช้วิธีการควบคุมอิทธิพลของตัวแปรร่วมอื่น ๆ ด้วยวิธี multivariable analysis โดยไม่ได้ประเมินอิทธิพลของ confounding by indicator ก่อนอาจทำให้สรุปผลการศึกษาผิดพลาด[1]

ในอดีตมักจะใช้วิธีจัดการทางด้านระเบียบวิธีวิจัย เพื่อควบคุมอิทธิพลของตัวแปรกวน ได้แก่ วิธีการ restriction หรือ การ matching แต่ก็ยังมีข้อจำกัด ที่วิธีการดังกล่าวสามารถควบคุมอิทธิพลของตัวแปรกวนได้เพียงหนึ่งหรือสองตัวแปรเท่านั้น จึงมักใช้วิธีจัดการทางสถิติร่วมด้วย ได้แก่ การวิเคราะห์แบบ stratification analysis หรือการวิเคราะห์ด้วย Multivariate analysis[6] แต่อย่างไรก็ตาม ยังคงพบว่ามีข้อจำกัดที่ต้องคำนึงถึงในกรณีที่มีจำนวนตัวแปรต่าง ๆ ที่เป็นทั้งตัวแปรร่วมและตัวแปรกวนจำนวนมาก ซึ่งหากต้องนำตัวแปรทั้งหมดเหล่านั้นเข้าไปอยู่ในสมการทำนายผล จะทำให้ความแม่นยำในการทำนายผลลดลง ซึ่งโดยทั่วไปการประมาณการว่าขนาดตัวอย่างมีความเพียงพอสำหรับใช้ในการวิเคราะห์ logistic regression model จะพิจารณาอย่างง่ายจากจำนวนตัวแปรที่อยู่ในโมเดลควรน้อยกว่าขนาดตัวอย่างในการศึกษานั้น ๆ อย่างน้อยที่สุดร้อยละ 10 เช่น ขนาดตัวอย่างในการศึกษานั้นมี 100 ตัวอย่าง ดังนั้นตัวแปรที่นำเข้าในโมเดลควรน้อยกว่า 10 ตัวแปร เป็นต้น[7]

ปัจจุบันมีวิธีการทางสถิติที่กล่าวขานและใช้กันเพิ่มมากขึ้นในรูปแบบการศึกษาเชิงสังเกต เนื่องจากงานวิจัยรูปแบบนี้มักจะมีตัวแปรกวนและตัวแปรร่วมหลายตัว การรวบตัวแปรร่วมและตัวแปรกวนรวมเป็นเพียงหนึ่งตัวแปรเรียกว่า คะแนนความโน้มเอียง หรือ propensity score และหลังจากนั้นจึงทำการประมาณค่าคะแนน propensity score ระหว่างกลุ่ม ซึ่งในบทความนี้จะได้สรุปแนวคิดในการใช้ propensity score และอธิบายวิธีการสร้าง propensity score และการนำ propensity score ไปใช้ในการควบคุมอิทธิพลของตัวแปรกวน[8,9]

แนวคิดเกี่ยวกับการวิเคราะห์คะแนนความโน้มเอียง (Propensity scores)

การวิเคราะห์คะแนนความโน้มเอียง (propensity scores) ใช้ครั้งแรกโดย Rosenbaum and Rubin[10] ในปี ค.ศ. 1983 โดยใช้การสรุปข้อมูลของตัวแปรร่วมหลาย ๆ ตัวแปรให้เป็น 1 ตัว เพื่อบอกโอกาสในการได้รับสิ่งแทรกแซงของกลุ่มตัวอย่างแต่ละคน โดยใช้คะแนนความน่าจะเป็น ที่คำนวณได้จากตัวแปรร่วมซึ่งเป็นลักษณะเฉพาะของแต่ละบุคคลนั้น เช่น โอกาสได้รับการรักษา ขึ้นอยู่กับอายุ เพศ ความรุนแรง ถ้าหากไม่เข้าเกณฑ์ที่กล่าวมาจะไม่ได้รับการรักษาด้วยวิธีนี้ เป็นต้น propensity score จึงเป็นทางเลือกในการประเมินโอกาสการได้รับการรักษาโดยอยู่บนเงื่อนไขของลักษณะพื้นฐานหรือลักษณะเฉพาะบุคคลของกลุ่มศึกษาดังกล่าว โดยพิจารณาจากความสมดุลของคะแนนความแนวโน้มที่คำนวณได้จากตัวแปรร่วมหลาย ๆ ตัว

จากการสืบค้นคำว่า “propensity score” จากฐานข้อมูล PubMed ในช่วงระยะเวลา 20 ปีที่ผ่านมาพบว่ามีบทความงานวิจัยที่ใช้วิธีวิเคราะห์ด้วย propensity score เพิ่มขึ้นรวดเร็วและต่อเนื่อง (รูปที่ 1)

คำจำกัดความทางสถิติคะแนนความโน้มเอียง

คะแนนความโน้มเอียง (Propensity scores) เป็นรูปแบบความน่าจะเป็นแบบมีเงื่อนไข (conditional probability) โดยอยู่บนเงื่อนไข 2 ประการ คือ 1. ผลของการให้การรักษาหรือผลของสิ่งทดลอง (intervention) ขึ้นอยู่กับลักษณะพื้นฐานของกลุ่มตัวอย่าง (baseline covariates) ซึ่งจะหมายรวมถึง “ตัวแปรกวนที่ไม่ได้ถูกวัดด้วย” และ 2. ทุก ๆ หน่วยของกลุ่มตัวอย่างจะต้องมีโอกาสในการรับการรักษาหรือสิ่งที่ทดลอง (โอกาสต้องไม่เป็นศูนย์)[11] และคะแนน propensity scores ต้องมีค่าอยู่ระหว่าง 0 ถึง 1 หรือเรียกว่า balancing score ใช้สำหรับดูการแจกแจงลักษณะของตัวแปรร่วมว่ามีลักษณะคล้ายกันหรือไม่ระหว่างกลุ่มที่ได้รับและไม่ได้รับการรักษาหรือสิ่งทดลอง สูตรคำนวณทางสถิติในการประมาณค่า propensity scores ตามที่ Rosenbaum and Rubin[10] ได้เสนอแนะไว้ ดังนี้

$$e(\mathbf{x}_i) = \Pr(z_i = 1 \mid \mathbf{x}_i)$$

โดยที่ กลุ่มตัวอย่างคนที่ *i,(i = 1,..., N)* จะได้รับการรักษาก็ต่อเมื่อมีตัวแปรร่วมคือ *Xi*

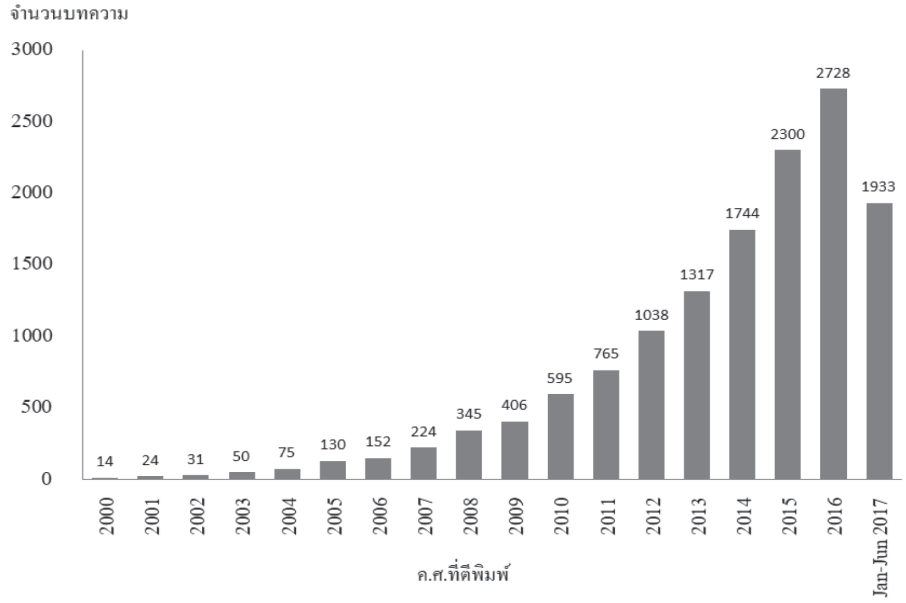
Z = การได้รับการรักษา หรือการได้รับสิ่งทดลอง, เป็นข้อมูล dichotomous data (0,1)

Zi = 1, ได้รับการรักษา

Zi = 0, ไม่ได้รับการรักษา

Xi = ตัวแปรร่วมต่างๆ ที่พบในกลุ่มตัวอย่างทำการศึกษา

เพื่อให้มองเห็นภาพที่ชัดเจนขึ้นเกี่ยวกับแนวคิดการวิเคราะห์ค่าคะแนนความโน้มเอียง จะขอยกตัวอย่างจากการศึกษาผลที่เกิดขึ้นเมื่อผู้ป่วยเบาหวานชนิดที่ 2 เกิดภาวะไตวาย โดยใช้ข้อมูลบางส่วนจากการศึกษาของ สิริมา มงคลสัมฤทธิ์ และคณะ[12] โดยกลุ่มตัวอย่างเป็นผู้ป่วยเบาหวาน จำนวน 4,112 ราย มี 1,590 รายที่มีภาวะไตวาย (เป็นกลุ่มศึกษา หรือ case) และอีก 2,522 รายไม่มีภาวะไตวาย (เป็นกลุ่มควบคุม หรือ control) ซึ่งในเชิงทฤษฎีมีหลายปัจจัยที่เกี่ยวข้องกับภาวะไตวายในผู้ป่วยเบาหวานชนิดที่ 2 ได้แก่ ระดับน้ำตาลในเลือด ความดันโลหิต เพศ อายุ ไขมันในเลือด การสูบบุหรี่ การดื่มแอลกอฮอล์ ภาวะอ้วน การรับประทานเค็ม เป็นต้น การจัดการตัวแปรที่มีอิทธิพลเหล่านี้สามารถใช้วิธีการคำนวณคะแนนความโน้มเอียงโดยใช้ตัวแปรที่กล่าวมาข้างต้นเป็นตัวทำนายด้วยโมเดลสมการถดถอย และคะแนนความน่าจะเป็นของการทำนายจะถูกนำไปจัดเป็นกลุ่มผู้ป่วยเบาหวานที่มีภาวะไตวาย และไม่มีภาวะไตวาย โดยอยู่บนพื้นฐานของตัวแปรที่ถูกนำเข้าไปวิเคราะห์ใน logistic regression model ถ้ากลุ่มตัวอย่างมีคะแนนความโน้มเอียงเท่ากัน สามารถบ่งบอกได้ว่ามีโอกาสเหมือนๆ กันที่จะถูกจัดอยู่ในกลุ่มที่มีภาวะไตวายหรือกลุ่มไม่มีไตวาย[1]



รูปที่ 1: จำนวนบทความที่ตีพิมพ์ในฐานข้อมูล PubMed

ขั้นตอนการคำนวณคะแนนความโน้มเอียง (Propensity score calculation)

การประมาณค่า propensity score ได้มาจาก logistic regression modelโดยที่ตัวแปรตาม (dependent variable) คือ การได้รับหรือไม่ได้รับ intervention หรือ treatment ส่วนตัวแปรต้น (independent variables) คือ ปัจจัยต่าง ๆ ที่เกี่ยวข้อง

ขอยกตัวอย่างการวิเคราะห์ผลของการใช้ยา RAS blocker เป็น intervention ในการป้องกันการเกิดโรคไตวาย (คือ dependent variable) ในผู้ป่วยเบาหวานชนิดที่ 2 ซึ่งจากการทบทวนวรรณกรรม พบว่า ปัจจัยที่เกี่ยวข้องกับการเกิดโรคไตวายมีหลายปัจจัย (independent variables) ได้แก่ ระดับน้ำตาลในเลือด ความดันโลหิต เพศ อายุ ไขมันในเลือด การสูบบุหรี่ การดื่มแอลกอฮอล์ ภาวะอ้วน เป็นต้น ในบทความนี้จะแสดงขั้นตอนการวิเคราะห์ propensity score ด้วยโปรแกรม STATA เวอร์ชั่น 14 ตามขั้นตอนดังนี้

1. วิเคราะห์สัดส่วนของ intervention ซึ่งในตัวอย่างนี้ คือ การใช้ยาและไม่ใช้ยา RAS blocker โดยกำหนดรหัสตัวแปร 0 หมายถึง เป็นกลุ่มที่ไม่ได้รับ RAS blocker และตัวแปร 1 หมายถึง เป็นกลุ่มที่ได้รับ RAS blocker (รูปที่ 2)

2. เลือกตัวแปรซึ่งเป็นตัวแปรพื้นฐานของกลุ่มตัวอย่าง ตัวแปรที่เป็นตัวแปรกวน และตัวแปรที่น่าจะเป็นตัวแปรที่ทำให้เกิด confounding by indicator และทำการสร้างโมเดลเริ่มต้น เพื่อคำนวณค่า propensity score ด้วย logistic regression model โดยกำหนดให้กลุ่มที่ได้รับการรักษา เป็นตัวแปรผลการศึกษา (outcome) และลักษณะของกลุ่มตัวอย่าง เช่น เพศ อายุ ความรุนแรงของโรค และตัวแปรอื่น ๆ เป็นตัวแปรทำนาย (predictor variables) ดังรูปที่ 3

```
**************************************************
Algorithm to estimate the propensity score
**************************************************

The treatment is drug
```

0 = No ACE/ARB 1= ACE/ARB	Freq.	Percent	Cum.
0	2,104	51.17	51.17
1	2,008	48.83	100.00
Total	4,112	100.00	

รูปที่ 2: สัดส่วนของการได้รับ Intervention

```
Estimation of the propensity score

Iteration 0:   log likelihood = -2613.7136
Iteration 1:   log likelihood = -2435.1552
Iteration 2:   log likelihood = -2434.2006
Iteration 3:   log likelihood = -2434.1997

Probit regression                                 Number of obs   =       3771
                                                  LR chi2(19)     =     359.03
                                                  Prob > chi2     =     0.0000
Log likelihood = -2434.1997                       Pseudo R2       =     0.0687
```

drug	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0062278	.0023228	2.68	0.007	.0016752	.0107804
sex	.2310197	.0565934	4.08	0.000	.1200986	.3419407
duration	.0033385	.0031465	1.06	0.289	-.0028285	.0095056
sbp_1	.0046467	.0012673	3.67	0.000	.0021629	.0071305
dbp_1	.0056724	.0024057	2.36	0.018	.0009572	.0103875
hba1c	-.0043455	.012622	-0.34	0.731	-.0290841	.020393
t_cholSI	-.0315418	.0455672	-0.69	0.489	-.120852	.0577683
tgSI	.0335489	.0248648	1.35	0.177	-.0151851	.0822829
hdlSI	-.0637372	.0719647	-0.89	0.376	-.2047854	.0773109
ldlSI	-.0215203	.0479103	-0.45	0.653	-.1154226	.0723821
crSI	-.0008418	.0003362	-2.50	0.012	-.0015007	-.0001829
hx_ihd	.0722343	.0832139	0.87	0.385	-.090862	.2353305
smoke	-.0437218	.0660218	-0.66	0.508	-.1731222	.0856786
alc	.0401011	.0671458	0.60	0.550	-.0915022	.1717044
insulin	.0651712	.0590823	1.10	0.270	-.050628	.1809705
drug_sugar	.44713	.078312	5.71	0.000	.2936413	.6006188
drug_HT	.2599269	.0454763	5.72	0.000	.170795	.3490587
drug_lipid	.1078443	.0444806	2.42	0.015	.020664	.1950246
nephro	.4723407	.0465752	10.14	0.000	.3810549	.5636264
_cons	-2.032317	.2753993	-7.38	0.000	-2.57209	-1.492545

รูปที่ 3: การคำนวณค่า Propensity score

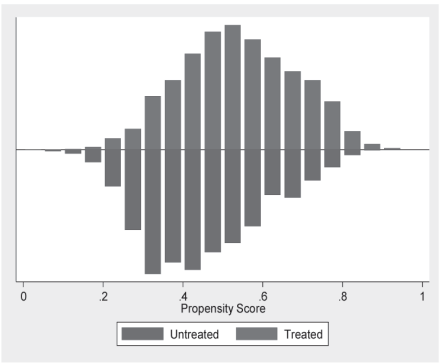
3. ประเมินความเท่ากันของคะแนนความน่าจะเป็น (propensity score) ที่คำนวณได้โดยจับคู่คะแนนความน่าจะเป็น ของกลุ่มที่ได้รับยาและไม่ได้รับยาว่าตรงกันหรือไม่ หากคะแนนความน่าจะเป็นของทั้งกลุ่มไม่มีคู่ที่สามารถจับคู่กันได้ ไม่จำเป็นต้องนำคะแนนความน่าจะเป็น ไปใช้ในการวิเคราะห์ในขั้นตอนถัดไป แต่หากการจับคู่คะแนนความน่าจะเป็น ของกลุ่มที่ได้รับยาและไม่ได้รับยาว่ามีลักษณะเท่า ๆ กัน แสดงว่าไม่มี confounding by indicator แต่หากการจับคู่คะแนนความน่าจะเป็นของกลุ่มที่ได้รับยาและไม่ได้รับยาว่ามีลักษณะไม่เท่ากัน แสดงว่าการได้รับการรักษาด้วยยาชนิดนั้นขึ้นอยู่กับลักษณะของผู้ป่วย ถ้าผู้ป่วยมีลักษณะตัวแปรพื้นฐานเหล่านี้ (ตัวแปรต้นต่าง ๆ ที่นำเข้าในโมเดล) จะมีแนวโน้มได้รับการรักษาด้วยยาชนิดนั้น ซึ่งเป็น confounding by indicator จำเป็นต้องนำค่า propensity score ไปใช้ในการวิเคราะห์ข้อมูลขั้นตอนต่อไป ซึ่งอาจวิเคราะห์แบบ matching ร่วมกับ propensity score หรือ regression ร่วมกับ propensity score เป็นต้น

จากรูปที่ 4 แปลผลได้ว่า คะแนนความโน้มเอียงของผู้ป่วยกลุ่มที่ได้รับ RAS blocker มีค่าเฉลี่ยเท่ากับ 0.62 ± 0.12 คะแนน และผู้ป่วยกลุ่มที่ไม่ได้รับ RAS blocker มีค่าเฉลี่ยคะแนนความโน้มเอียงของเท่ากับ 0.42 ± 0.11 คะแนน ซึ่งคะแนนความโน้มเอียงของทั้งสองกลุ่มมีความแตกต่างกัน (p value <0.001) ในกรณีนี้ควรต้องนำตัวแปรคะแนนความโน้มเอียงไปใช้วิเคราะห์ข้อมูลต่อร่วมกับวิธีการอื่น ๆ ซึ่งได้กล่าวในหัวข้อถัดไป

ซึ่งข้อดีของการใช้คะแนนความโน้มเอียงจากตัวอย่างดังกล่าวจะทำให้สามารถประเมินได้ว่ามี confounding by indicator หรือไม่ และเป็นการสรุปรวบตัวแปรร่วมหลาย ๆ ตัวให้เหลือเพียง 1 ตัว แต่ข้อด้อยของการวิเคราะห์คะแนนความโน้มเอียง คือ ตัวแปรร่วมบางตัวหรือตัวแปรกวนที่ไม่มีการเก็บเข้าในการศึกษาจะไม่สามารถประเมิน confounding by indicator ได้

การใช้คะแนนความโน้มเอียงร่วมกับวิธีอื่นๆ

การใช้ค่า propensity score เพื่อขจัดอิทธิพลของตัวแปรกวนเมื่อต้องการศึกษาผลของการรักษาสามารถทำได้หลายแบบ เช่น propensity score matching (PSM), stratification on propensity score, การถ่วงน้ำหนักด้วย propensity score (weighting propensity score) และการวิเคราะห์แบบ multivariable analysis ร่วมกับ propensity score[1]



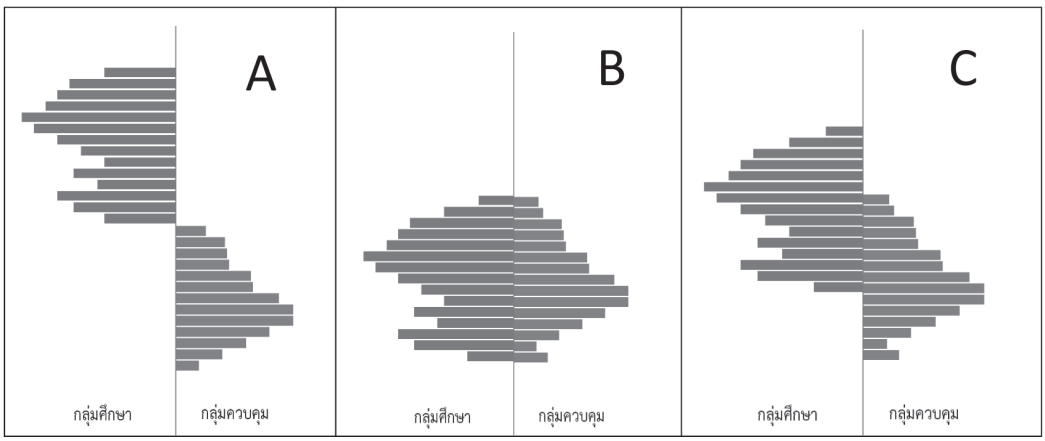

```
. ttest propensity, by(neph)

Two-sample t test with equal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |   2,303    .4171909    .0023264    .1116452    .4126287    .421753
       1 |   1,468    .6168983    .0031625    .1211705    .6106947    .6231018
---------+--------------------------------------------------------------------
combined |   3,771    .4949343    .0024594    .1510255    .4901125    .4997561
---------+--------------------------------------------------------------------
    diff |           -.1997074    .0038557                -.2072667   -.192148
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t = -51.7960
Ho: diff = 0                                     degrees of freedom =     3769

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

รูปที่ 4: การประเมินการแจกแจง propensity score ระหว่างกลุ่มที่ได้รับและไม่ได้รับ RAS blocker



รูปที่ 5: รูปแบบการประเมินการแจกแจง propensity score แบบต่าง ๆ

1. Propensity Score Matching (PSM)

Propensity score matching เป็นวิธีการทางสถิติที่ใช้เพื่อการจับคู่ของกลุ่มศึกษา (case) และกลุ่มควบคุม (control) โดยใช้คะแนน propensity score ของกลุ่มศึกษาจับคู่กับคะแนน propensity score ของกลุ่มควบคุม โดยพิจารณาจากคะแนน propensity score ที่มีค่าเท่ากันหรือใกล้เคียงจับคู่กัน ซึ่งการจับคู่แบบนี้สามารถช่วยทำให้การอ้างเหตุผลเชิงสาเหตุและผลมีความชัดเจนขึ้นสำหรับรูปแบบการศึกษากึ่งทดลอง (quasi experiment) และรูปแบบการศึกษาเชิงสังเกต (observational study) ในการลดอคติจากการคัดเลือก (selection bias) ซึ่งในปัจจุบันมีการใช้ propensity score matching จำนวนมาก[13-16] ซึ่งโดยปกติจะพบวิธีการจับคู่แบบ 1:1 matching มากที่สุด ซึ่งเป็นวิธีการดั้งเดิมแต่มีข้อจำกัดในกรณีที่กลุ่มตัวอย่างมีขนาดเล็กหรือในกรณีที่หากลุ่มศึกษายาก ก็จะใช้วิธีการจับคู่กลุ่มศึกษาและกลุ่มควบคุมแบบ 1:2, 1:3 หรือ 1:4 matching แต่ก็พบว่ามีการจับคู่ propensity score แบบ pair matching โดยคะแนนความโน้มเอียงระหว่างกลุ่มศึกษาและกลุ่มควบคุมต้องเท่ากัน การจับคู่แบบ nearest-neighbor matching หรือ greedy matching โดยคู่ระหว่างกลุ่มศึกษาและกลุ่มควบคุมต้องมีค่า propensity score ที่ใกล้เคียงกันมากที่สุด การจับคู่แบบ optimal matching และการจับคู่แบบ caliper matching มีค่า ความกว้างประมาณ ±0.01 หรือ ±.025 ซึ่งหลังการ matching จะทำการเปรียบเทียบคะแนนความโน้มเอียงระหว่างกลุ่มศึกษาและกลุ่มควบคุมด้วยสถิติ independent t-test หรือ paired t-test

2. Stratification on propensity score

เป็นวิธีการแบ่งกลุ่มตัวอย่างทั้งหมดออกเป็นระดับชั้นโดยใช้การอ้างอิงของค่า propensity score ที่ประมาณค่าไว้โดยเรียงตามลำดับคะแนนของความน่าจะเป็น และทำการวิเคราะห์แยกภายในแต่ละระดับชั้น หลังจากนั้นจะทำการวิเคราะห์ความแตกต่างของคะแนนความโน้มเอียงด้วยวิธี two-group ANOVA ซึ่งเป็นข้อดีของการใช้คะแนนความโน้มเอียงเพียงตัวเดียวในการควบคุมความแตกต่าง ซึ่งการควบคุมตัวแปรกวนด้วย stratification จะสามารถทำได้เพียงหนึ่งหรือสองตัวแปรเท่านั้น[17-20]

3. Weighting on propensity score

เป็นการวิเคราะห์จากความผกผันของการประมาณคะแนน propensity score หรือ ค่า inverse probability weighting (IPW) ซึ่งเป็นการทำ weighting on propensity score จะใช้บ่อยมากในการวิเคราะห์การรอดชีพ (survival analysis) หรือการศึกษาข้อมูลของประชากร เช่น ทางด้านระบาดวิทยาชุมชน ทางด้านสาธารณสุข หรือ ทางด้านเศรษฐศาสตร์[18,19]

4. Multivariable analysis with propensity score

เป็นวิธีการที่ง่าย โดยการนำค่า propensity score ที่คำนวณได้ ไปเข้าโมเดลการวิเคราะห์สมการถดถอย เสมือนว่าตัวแปรคะแนนความโน้มเอียงเป็นตัวแปรร่วมอีกหนึ่งตัวในโมเดลเพื่อการวิเคราะห์ประมาณค่าผลการศึกษา แต่ประเด็นที่นักวิจัยต้องคำนึงในการแปลผลการศึกษา คือ การแปลผลอยู่บนพื้นฐานของตัวแปรที่นำเข้าไปใช้ในการคำนวณคะแนนความโน้มเอียงเท่านั้น การแปลผลที่นอกเหนือจากตัวแปรที่ไม่ได้มีอยู่ในการศึกษาจะทำให้เกิดความคลาดเคลื่อนของผลการวิจัย[21-24]

กรณีที่ไม่สามารถใช้คะแนนความโน้มเอียง

คะแนนความโน้มเอียง เป็นเครื่องมือที่ช่วยสนับสนุนการศึกษาเชิงเหตุเชิงผลภายใต้ข้อจำกัดที่เคยมีมาก่อนหน้านี้ แต่อย่างไรก็ตาม มีบางประเด็นที่ไม่สามารถนำคะแนนความโน้มเอียงไปใช้ได้[1] ดังนี้

1. คะแนนความโน้มเอียง เป็นเพียงการป้องกันอคติที่เกิดขึ้นจากเฉพาะตัวแปรที่ถูกนำไปใช้ในการคำนวณเท่านั้น หากมีตัวแปรกวนสำคัญที่ไม่ได้ถูกนำไปใช้ในการวิเคราะห์หาค่าคะแนนความโน้มเอียงด้วย ก็จะทำให้ผลการศึกษาเกิดความผิดพลาดได้

2. เทคนิคที่เรียกว่า “Propensity score calibration” เป็นวิธีการที่ถูกพัฒนาขึ้นมาเพื่อใช้ในการปรับอคติในกรณีที่คะแนนความโน้มเอียงไม่มีการวัดตัวแปรกวน ซึ่งจำเป็นต้องมีอีกหนึ่งการศึกษาที่กลุ่มประชากรเหมือนกัน แต่ขาดการเก็บข้อมูลตัวแปรกวน ซึ่งในกรณีนี้มีความเป็นไปได้ยากที่นักวิจัยจะเลือกใช้วิธีการนี้

3. คะแนนความโน้มเอียงจะสามารถนำไปใช้ได้ในกรณีที่มีการกระจายของคะแนนความโน้มเอียงเท่าๆ กัน หรือใกล้เคียงกันระหว่างกลุ่มศึกษาและกลุ่มควบคุม ซึ่งหากไม่เป็นไปตามนี้แล้วคะแนนความโน้มเอียงจะไม่สามารถจัดการอคติของผลการศึกษาได้

4. คะแนนความโน้มเอียงถือว่าเป็นตัวแปรตามตัวหนึ่งของกลุ่มตัวอย่าง และคะแนนนี้ไม่สามารถนำไปใช้โดยทั่วไปได้ เนื่องจากคะแนนความโน้มเอียงจะมีการเปลี่ยนแปลงไปตามกลุ่มตัวอย่างที่นำมาวิเคราะห์ และขึ้นอยู่กับจำนวนตัวแปรที่นำไปใช้ในการคำนวณคะแนนความโน้มเอียง

5. คะแนนความโน้มเอียงเป็นการขจัดอคติในการศึกษารูปแบบเชิงสังเกต และช่วยในการอธิบายเชิงสาเหตุ ซึ่งประเด็นสำคัญที่ต้องให้ความใส่ใจ คือ วิธีการนี้ไม่ได้ช่วยในเรื่องของการแก้ปัญหาของผลการศึกษาที่ไม่มีนัยสำคัญ (non-significant) และการวิจัยที่ออกแบบการศึกษาไม่ดี แต่ในทางตรงกันข้ามคะแนนความโน้มเอียง จะช่วยในกรณีที่การศึกษาเรื่องนั้นๆ มีการออกแบบการศึกษามาเป็นอย่างดี และมีการวางแผนการวิจัยที่มีคุณภาพแต่ไม่สามารถใช้วิธีการสุ่มแบบ random assignment ได้

สรุป

Propensity score มีประโยชน์ในการใช้ควบคุมตัวแปรกวนและตัวแปรร่วมที่มีขนาดใหญ่และมีการเกิดเหตุการณ์ที่พบน้อย ซึ่งจะช่วยประหยัดเวลาและงบประมาณ แต่ propensity score ก็มีข้อจำกัดบางประการ คือ สามารถควบคุมได้เฉพาะปัจจัยที่เป็นตัวแปรร่วมที่เก็บมาในการศึกษา หากตัวแปรใดที่เป็นตัวแปรกวนที่มีอิทธิพลแต่ไม่ได้ถูกเก็บข้อมูลมาใช้เพื่อการวิเคราะห์ propensity score จะไม่สามารถปรับค่าอิทธิพลของตัวแปรเหล่านั้นได้ และวิธีการใช้ propensity score จะดีในกรณีที่กลุ่มตัวอย่างมีขนาดใหญ่พอที่จะทำให้เห็นการแจกแจงของตัวแปรร่วม มีโอกาสเกิดอคติเนื่องจากการเลือกหรือไม่เลือกตัวแปรเข้าไปสร้างคะแนน propensity score

เอกสารอ้างอิง

1. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate behavioral research. 2011;46(3): 399-424.
2. Olthof DC, Joosse P, Bossuyt PM, de Rooij PP, Leenen LP, Wendt KW, et al. Observation Versus Embolization in Patients with Blunt Splenic Injury After Trauma: A Propensity Score Analysis. World J Surg. 2016;40(5):1264-71.
3. Hur M, Koo CH, Lee HC, Park SK, Kim M, Kim WH, et al. Preoperative aspirin use and acute kidney injury after cardiac surgery: A propensity-score matched observational study. PLoS One. 2017;12(5):e0177201.
4. Trojano M, Pellegrini F, Paolicelli D, Fuiani A, Di Renzo V. observational studies: propensity score analysis of non-randomized data. Int MS J / MS Forum. 2009;16(3):90-7.
5. Wang Z. Propensity score methods to adjust for confounding in assessing treatment effects: bias and precision. Internet J Epidemiol. 2008;7(2):1-7.
6. Mongkolsomlit S, Patumanond J, Rawdaree P. How to detect and handle confounding factors. Vajira Med J. 2010;54:223-35.
7. Rossi RJ. Applied Biostatistics for the health sciences: John Wiley & Sons, Inc; 1996.
8. Fitzmaurice G. Confounding: propensity score adjustment. Nutrition. 2006;22 (11-12):1214-6.
9. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. 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. J Clin Epidemiol. 2006;59(5):437-47.

10. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983;70(1):41-55.
11. Marshall MJ, Paul RR. Invited commentary: Propensity score. Am J Epidemiol. 1999;150: 327-33.
12. Mongkolsomlit S, Rawdaree P, Komoltri C, Tawichasri C, Patumanond J. Effect of Angiotensin-Converting Enzyme Inhibitors and/or Angiotensin Receptor Blockers on the Prevention of Death in Patients with Type 2 Diabetes and Undetermined Nephropathy : Five-Year Survival Data. J Diabetes Metab. 2012;3:188.
13. Justus J. Randolph, Falbe K, Manuel AK, Joseph L. Balloun. A Step-by- Step Guide to Propensity Score Matching in R. Practical Assessment, Research & Evaluation. 2014;19(18).
14. Duhamel A, Labreuche J, Gronnier C, Mariette C. Statistical Tools for Propensity Score Matching. Ann Surg. 2017;265(6):E79-E80.
15. Gray E, Pasta DJ, Norris S, O'Leary A, Irish Hepatitis CO, Research N. Effectiveness of triple therapy with direct-acting antivirals for hepatitis C genotype 1 infection: application of propensity score matching in a national HCV treatment registry. BMC Health Serv Res. 2017;17(1):288.
16. Yu HS, Hwang JE, Chung HS, Cho YH, Kim MS, Hwang EC, et al. Is preoperative chronic kidney disease status associated with oncologic outcomes in upper urinary tract urothelial carcinoma? a multicenter propensity score-matched analysis. Oncotarget. 2017;8(39): 66540-9.
17. Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF. A Propensity-score-based Fine Stratification Approach for Confounding Adjustment When Exposure Is Infrequent. Epidemiology. 2017;28(2):249-57.
18. Linden A, Adams J, Roberts N. Using propensity scores to construct comparable control groups for disease management program evaluation. Disease Management and Health Outcomes. 2005;13(2):107-15.
19. Linden A. Improving causal inference with a doubly robust estimator that combines propensity score stratification and weighting. J Eval Clin Pract. 2017;23(4):697-702.
20. Senn S, Graf E, Caputo A. Stratification for the propensity score compared with linear regression techniques to assess the effect of treatment or exposure. Stat Med. 2007;26(30):5529-44.
21. Bohn J, Eddings W, Schneeweiss S. Conducting Privacy-Preserving Multivariable Propensity Score Analysis When Patient Covariate Information Is Stored in Separate Locations. Am J Epidemiol. 2017;185(6):501-10.
22. Anderson EJ, Carosone-Link P, Yogev R, Yi J, Simoes EAF. Effectiveness of Palivizumab in High-risk Infants and Children: A Propensity Score Weighted Regression Analysis. Pediatr Infect Dis J. 2017;36(8):699-704.
23. Caron C, Wasser T, Eisenberg D. The Use of Propensity Score Matching Does Not Protect Against Regression Artifacts (Regression Towards The Mean). Value Health. 2015;18(7):A721.
24. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol. 2010;63(8):826-33.