Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

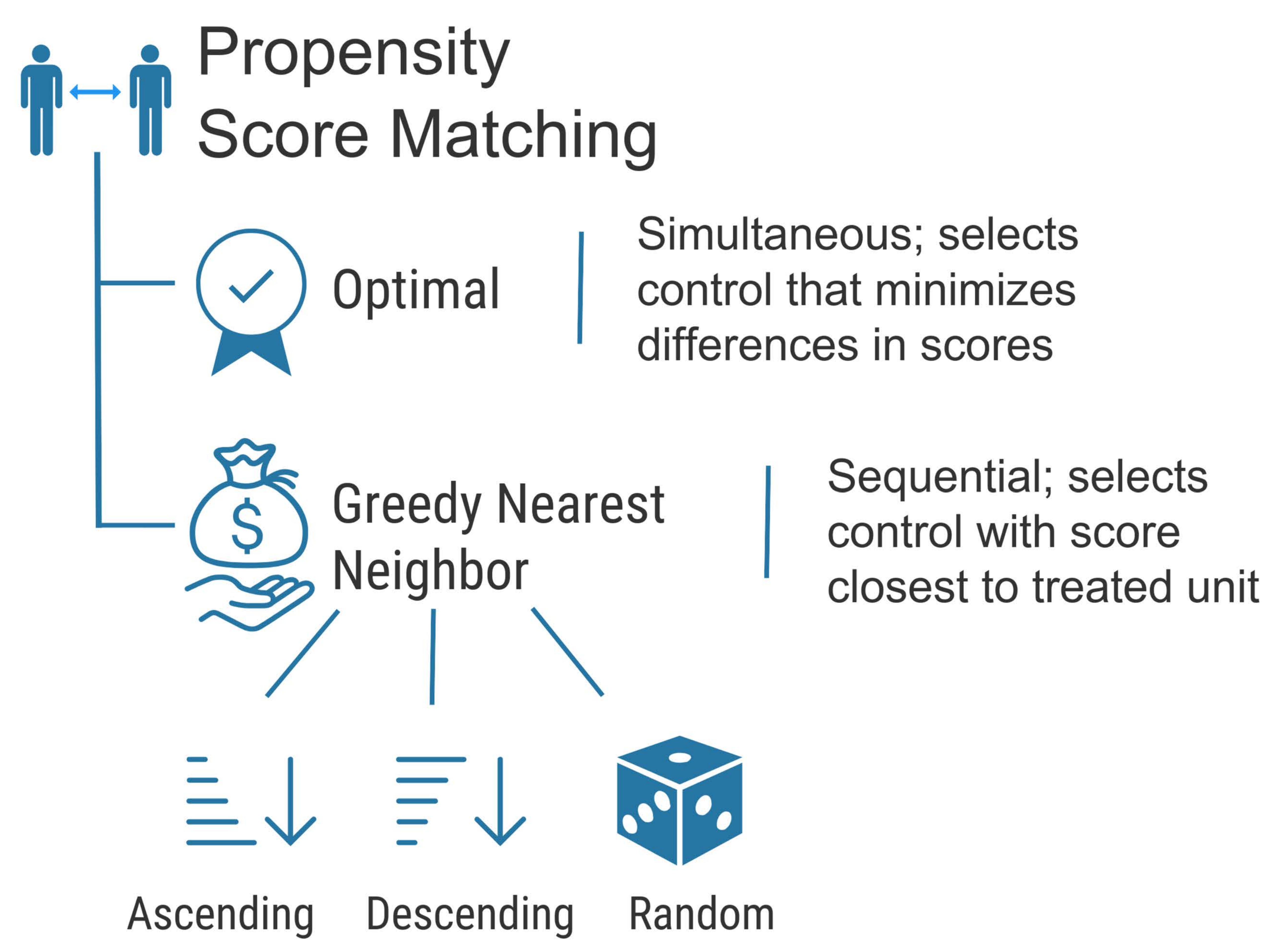
Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

First Author
@BN_Hand

Introduction

- Propensity scores are commonly used to reduce bias in observational studies
- PROC PSMATCH offers a number of methods for using propensity scores, eliminating the need for macros
- Proportion of matched cases depends on matching strategy, case to control ratio, and the number of variables.



Objective

To demonstrate the influence of data features and PROC PSMATCH settings on 1:1 propensity score matching of treated and control units without replacement.

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

First Author
@BN_Hand

Methods

Variation in Case to Control Ratio

CMS Data

2013-2015
Inpatient Standard
Analytic Files

↓

Patient Pool

Pancreatomy patients who were (cases) and were not (controls) discharged to skilled nursing facilities

Cases (n=3k) ↓ | Controls (n=17k) ↓

↓

Analytical Datasets

SRS of controls using PROC SURVEY SELECT, bootstrapped x1,000

Case n	Ratio	Control n
3k	1:1	3k
3k	1:2	6k
3k	1:3	9k
3k	1:4	12k
3k	1:5	15k

Variation in Number of Variables

# of Vars	Variables
5	Sex | Race | Age | Charlson Index | LOS
10	+ 5 Comorbidities
15	+ 5 Comorbidities
20	+ 5 Comorbidities
25	+ 5 Comorbidities

- Macro MatchyMatchy
- Series of propensity score models as a function of:
  1. Case to control ratio
  2. Number of variables in the model
  3. Matching strategy

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

First Author

@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	Matching strategy	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	70 (67.6, 72.6)	68.7 (66.3, 71.2)	68 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4,70.4)
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99)	97.5 (95.7, 98.9)	97.4 (95.8,98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	Optimal	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	99.6 (99.1, 99.9)	99.5 (98.6, 99.9)	99.4 (98.5, 99.9)	99.5 (98.5, 99.9)	99.4 (98.4,99.9)
	Random	97.6 (95.9, 99)	96.9 (95.3, 98.3)	96.7 (95, 98.3)	96.8 (95.1, 98.3)	96.7 (95, 98.2)
1:5	Optimal	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	99.9 (99.6, 100)	99.8 (99.4, 100)	99.8 (99.5, 100)	99.8 (99.5, 100)	99.9 (99.5, 100)
	Random	99.3 (98.3, 99.9)	99 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6,99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

Introduction
Methods
Results
Conclusion

First Author
@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	**Matching strategy**	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	**70 (67.6, 72.6)**	**68.7 (66.3, 71.2)**	**68 (65.5, 70.7)**	**67.9 (65.4, 70.5)**	**67.9 (65.4,70.4)**
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	**91.7 (89.6, 93.7)**	**90.4 (88.3, 92.5)**	**90.8 (88.6, 92.8)**	**90.4 (88.2, 92.5)**	**90.2 (88, 92.3)**
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	**97.9 (96.2, 99.5)**	**97.6 (95.8, 99.1)**	**97.4 (95.7, 99)**	**97.5 (95.7, 98.9)**	**97.4 (95.8,98.9)**
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	Optimal	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	**99.6 (99.1, 99.9)**	**99.5 (98.6, 99.9)**	**99.4 (98.5, 99.9)**	**99.5 (98.5, 99.9)**	**99.4 (98.4,99.9)**
	Random	97.6 (95.9, 99)	96.9 (95.3, 98.3)	96.7 (95, 98.3)	96.8 (95.1, 98.3)	96.7 (95, 98.2)
1:5	Optimal	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	**99.9 (99.6, 100)**	**99.8 (99.4, 100)**	**99.8 (99.5, 100)**	**99.8 (99.5, 100)**	**99.9 (99.5, 100)**
	Random	99.3 (98.3, 99.9)	99 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6,99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

Introduction
Methods
Results
Conclusion

First Author
@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	**Matching strategy**	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	70 (67.6, 72.6)	68.7 (66.3, 71.2)	68 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4,70.4)
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99)	97.5 (95.7, 98.9)	97.4 (95.8,98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	**Optimal**	**99.9 (99.8, 100)**	**99.9 (99.8, 100)**	**99.9 (99.8, 100)**	**99.9 (99.8, 100)**	**99.9 (99.7, 100)**
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	**99.6 (99.1, 99.9)**	**99.5 (98.6, 99.9)**	**99.4 (98.5, 99.9)**	**99.5 (98.5, 99.9)**	**99.4 (98.4,99.9)**
	Random	97.6 (95.9, 99)	96.9 (95.3, 98.3)	96.7 (95, 98.3)	96.8 (95.1, 98.3)	96.7 (95, 98.2)
1:5	**Optimal**	**100 (99.9, 100)**	**100 (99.9, 100)**	**100 (99.9, 100)**	**100 (99.9, 100)**	**100 (99.9, 100)**
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	**99.9 (99.6, 100)**	**99.8 (99.4, 100)**	**99.8 (99.5, 100)**	**99.8 (99.5, 100)**	**99.9 (99.5, 100)**
	Random	99.3 (98.3, 99.9)	99 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6,99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

First Author
@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	Matching strategy	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	70 (67.6, 72.6)	68.7 (66.3, 71.2)	68 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4,70.4)
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99)	97.5 (95.7, 98.9)	97.4 (95.8,98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	Optimal	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	99.6 (99.1, 99.9)	99.5 (98.6, 99.9)	99.4 (98.5, 99.9)	99.5 (98.5, 99.9)	99.4 (98.4,99.9)
	Random	97.6 (95.9, 99)	96.9 (95.3, 98.3)	96.7 (95, 98.3)	96.8 (95.1, 98.3)	96.7 (95, 98.2)
1:5	Optimal	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	99.9 (99.6, 100)	99.8 (99.4, 100)	99.8 (99.5, 100)	99.8 (99.5, 100)	99.9 (99.5, 100)
	Random	99.3 (98.3, 99.9)	99 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6,99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

THE OHIO STATE UNIVERSITY

Introduction
Methods
Results
Conclusion

First Author
@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	Matching strategy	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	70 (67.6, 72.6)	68.7 (66.3, 71.2)	68 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4,70.4)
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99)	97.5 (95.7, 98.9)	97.4 (95.8,98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	**Optimal**	**99.9 (99.8, 100)**	**99.9 (99.8, 100)**	**99.9 (99.8, 100)**	**99.9 (99.8, 100)**	**99.9 (99.7, 100)**
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	**99.6 (99.1, 99.9)**	**99.5 (98.6, 99.9)**	**99.4 (98.5, 99.9)**	**99.5 (98.5, 99.9)**	**99.4 (98.4,99.9)**
	Random	**97.6 (95.9, 99)**	**96.9 (95.3, 98.3)**	**96.7 (95, 98.3)**	**96.8 (95.1, 98.3)**	**96.7 (95, 98.2)**
1:5	**Optimal**	**100 (99.9, 100)**	**100 (99.9, 100)**	**100 (99.9, 100)**	**100 (99.9, 100)**	**100 (99.9, 100)**
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	**99.9 (99.6, 100)**	**99.8 (99.4, 100)**	**99.8 (99.5, 100)**	**99.8 (99.5, 100)**	**99.9 (99.5, 100)**
	Random	**99.3 (98.3, 99.9)**	**99 (97.9, 99.8)**	**98.9 (97.8, 99.7)**	**98.9 (97.7, 99.8)**	**98.8 (97.6,99.7)**

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

THE OHIO STATE UNIVERSITY

Introduction
Methods
Results
Conclusion

First Author
@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	Matching strategy	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	70 (67.6, 72.6)	68.7 (66.3, 71.2)	68 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4,70.4)
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99)	97.5 (95.7, 98.9)	97.4 (95.8,98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	Optimal	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	99.6 (99.1, 99.9)	99.5 (98.6, 99.9)	99.4 (98.5, 99.9)	99.5 (98.5, 99.9)	99.4 (98.4,99.9)
	Random	97.6 (95.9, 99)	96.9 (95.3, 98.3)	96.7 (95, 98.3)	96.8 (95.1, 98.3)	96.7 (95, 98.2)
1:5	Optimal	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	99.9 (99.6, 100)	99.8 (99.4, 100)	99.8 (99.5, 100)	99.8 (99.5, 100)	99.9 (99.5, 100)
	Random	99.3 (98.3, 99.9)	99 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6,99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

THE OHIO STATE UNIVERSITY

First Author
@BN_Hand

Results

- Results reported as mean % of cases matched (95% confidence interval)

Cases:Controls (input data)	Matching strategy	Number of Variables: 5	10	15	20	25
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61)	58.4 (55.9, 60.8)	58.5 (56, 60.8)
	Descending	70 (67.6, 72.6)	68.7 (66.3, 71.2)	68 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4,70.4)
	Random	62.5 (60.2, 64.9)	61 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9,62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83 (80.8, 85.1)	82.7 (80.5, 85)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93)	91 (88.7, 93)	90.9 (88.6, 93.1)	90.8 (88.5,92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99)	97.5 (95.7, 98.9)	97.4 (95.8,98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94)	92.1 (90.1, 94.1)	92 (90, 93.9)
1:4	Optimal	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96 (94, 97.7)	95.9 (94, 97.7)
	Descending	99.6 (99.1, 99.9)	99.5 (98.6, 99.9)	99.4 (98.5, 99.9)	99.5 (98.5, 99.9)	99.4 (98.4,99.9)
	Random	97.6 (95.9, 99)	96.9 (95.3, 98.3)	96.7 (95, 98.3)	96.8 (95.1, 98.3)	96.7 (95, 98.2)
1:5	Optimal	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9,99.8)
	Descending	99.9 (99.6, 100)	99.8 (99.4, 100)	99.8 (99.5, 100)	99.8 (99.5, 100)	99.9 (99.5, 100)
	Random	99.3 (98.3, 99.9)	99 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6,99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

Matchy-matchy: Role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD, MPH

The Ohio State University

Introduction
Methods
Results
Conclusion

First Author
@BN_Hand

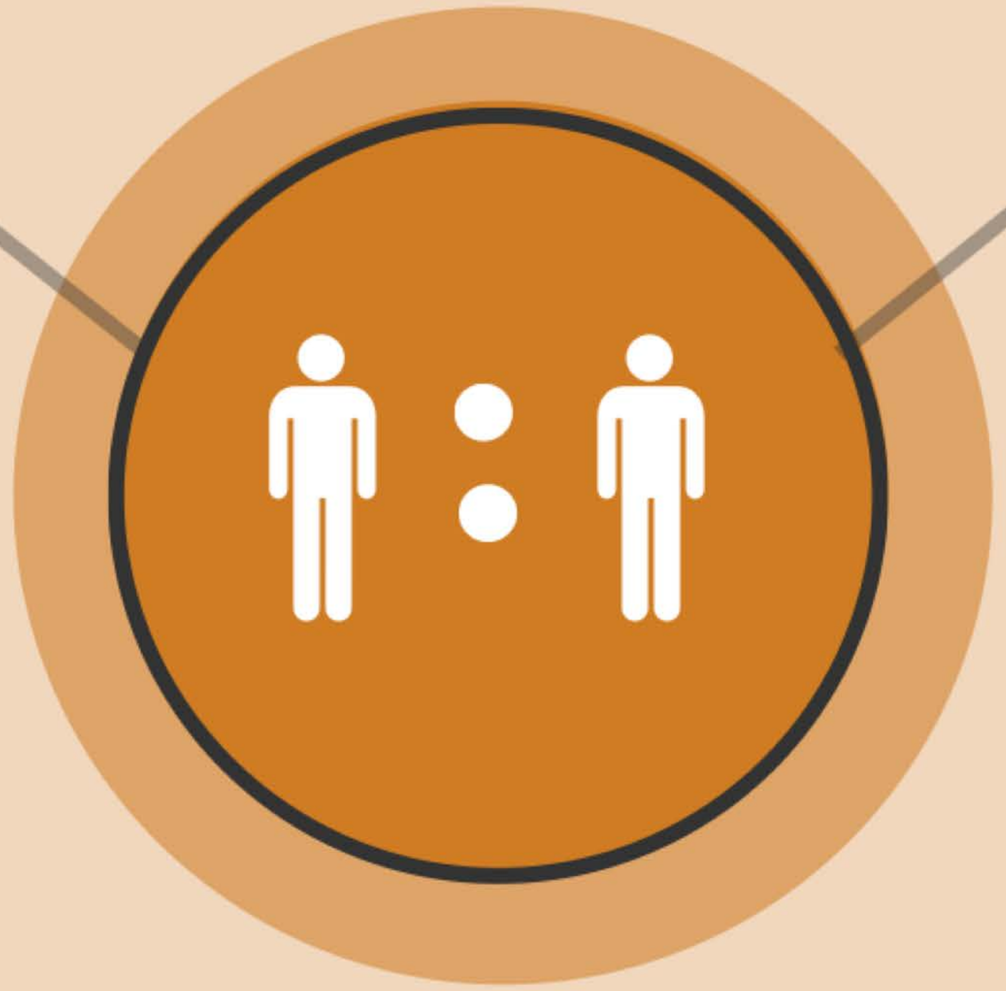
Conclusion

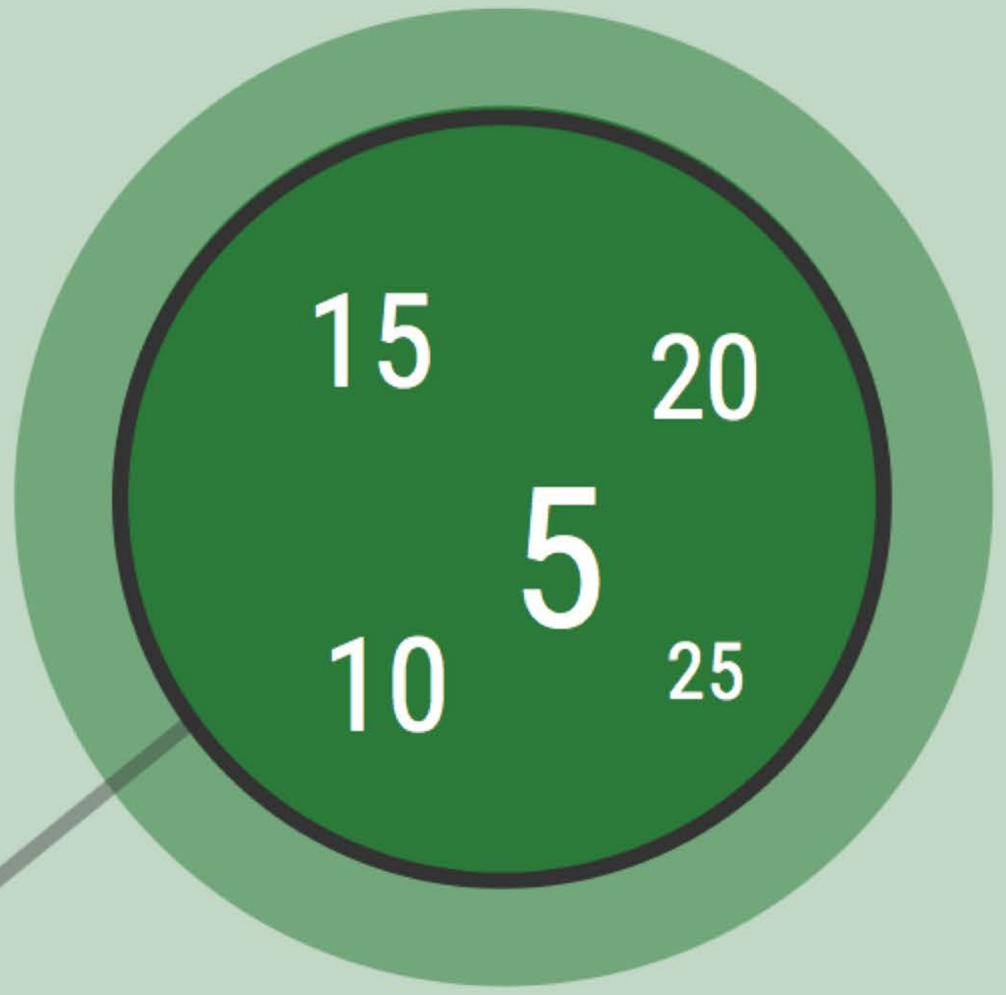
Strategy

- Greedy-descending out-performed greedy-ascending & greedy-random
- If case:control = 1:4 or greater, greedy-descending and optimal strategies performed similarly
- Optimal strategy more computationally intensive than greedy descending

Case : Control

- Increasing case to control ratio improved the proportion of matched cases.
- As the number of controls relative to the number of cases becomes excessively large, diminishing returns on proportion of matched cases





Number of Variables

- As the number of variables increases, the proportion of matched cases generally decreases
- With a large number of controls, adding variables had less of an effect on the proportion of matched cases

Paper 3620-2019

Matchy-matchy: role of strategy, case to control ratio, and number of variables in PROC PSMATCH

Brittany N. Hand, PhD, J. Madison Hyer, M.S., Rittal Mehta, MPH, Timothy Pawlik, MD, PhD

ABSTRACT

Propensity score matching is commonly used in observational studies to adjust data for increased validity in exposure variable effect estimation. PROC PSMATCH offers a number of methods for using propensity scores to this end, eliminating the need for macros. This paper will highlight the use of PROC PSMATCH to perform 1:1 propensity score matching of treated and control units without replacement. As the proportion of matched cases is influenced by features of the data and PROC PSMATCH settings, this paper provides an example to illustrate these effects. The proportion of matched units are evaluated as a function of matching strategy (i.e., greedy vs. optimal), ratio of cases to controls in the original dataset, and the number of variables in the model.

INTRODUCTION

Propensity score matching is commonly used in observational studies to retrospectively approximate randomization by reducing the effects of confounding. This technique has become increasingly used in health services research to examine the efficacy of clinical interventions and make inferences about policy change using large, administrative billing databases.

PROC PSMATCH offers two types of strategies (Yuan, Yung & Stokes) for 1:1 matching without replacement -- greedy nearest neighbor and optimal matching. The greedy matching strategy is further classified as ascending, descending, or random based on the ordering of the propensity scores generated by the model. Table 1 provides a brief description of these strategies.

Table 1: PROC PSMATCH Matching Strategies

Matching Strategy	**Description**
Optimal (fixed ratio)	Simultaneously selects the control units that minimize the total absolute difference in propensity scores across all matches
Greedy nearest neighbor	Sequentially selects the control unit whose propensity score is closest to the treated unit
Ascending	Orders cases in ascending order of propensity score
Descending (default)	Orders cases in descending order of propensity score
Random	Orders cases in random order of the propensity score

The purpose of this paper is to demonstrate the results of different matching strategies offered in PROC PSMATCH as a function of the: 1) ratio of cases to controls in the original data and 2) number of variables in the model.

DATA SETUP

Data were derived from the 2013-2015 Inpatient Standard Analytic Files (SAFs) from the Centers of Medicare and Medicaid Services. In this example, a researcher wants to compare outcomes of patients who underwent pancreatomy who were (cases) and were not (controls) discharged to a skilled nursing facility after surgery. The control units will be 1:1 matched to cases based on age, sex, race/ethnicity, Charlson comorbidity index, and presence of perioperative complications.

Pool of eligible units: An original dataset consisting of n=3,000 cases and n=17,000 controls was used in this study. To study the effects of the ratio of cases to controls in the input dataset, the following 5 datasets were generated using PROC SURVEY SELECT:

1) oneone = 1:1 ratio of cases to controls
2) onetwo = 1:2 ratio of cases to controls
3) onethree = 1:3 ratio of cases to controls
4) onefour = 1:4 ratio of cases to controls
5) onefive= 1:5 ratio of cases to controls

All datasets contained 100% of cases, and a random selection of either 3k, 6k, 9k, 12k, 15k controls. The variable GROUP was used to indicate cases and controls. The data included a 9-digit numeric identifier (ID), age (AGE), sex (MALE), length of stay (LOS), race/ethnicity (RACE), Charlson Comorbidity index (CHARLSSCORE), and dichotomous indicators for the presence of 20 comorbid medical conditions (CMC1-CMC20).

DATA ANALYSIS

To estimate the expected proportion and the variability of matched cases and controls as a function of matching strategy, number of variables, and case to control ratio, propensity score matching was performed on 1,000 bootstrapped samples. A SAS Macro was written to iteratively perform the propensity score matching procedure and output relevant information. Additionally, other macros were written to compile the results and construct tables for presenting results. Below is an example of the macro that performs the propensity score matching procedure (specifically, the greedy ascending matching strategy). The unabridged version of the macros are presented at the end of the document in Supplemental Material.

```
proc psmatch data=%scan(&data., &i)_c region=allobs;
      where Replicate = &iteration.;
      class group &Vars_Exact.;
      psmodel group(Treated="Case") = &Vars. &Vars_Exact.;
      match method=greedy(k=1 order = ascending) stat=lps;
      assess lps var=(&Vars.) / weight=none plots=none;
      output out(obs=match)=&Vars_num._var_asc_&i. lps=_lps matchid=_MatchID;
run;
```

`%scan(&data., &i)_c:` Dataset
`&iteration.:` Iteration i of the 1,000 iterations
`&Vars_Exact.:` Variables to be matched exactly (was identical for all experimental conditions)
`&Vars.:` Variable list that was added to for assessing varying number of variables

RESULTS

The results are summarized in Table 2 and described below.

Matching strategy: Table 2 shows that the greedy-descending strategy consistently out-performed the greedy-ascending and greedy-random strategies. When the ratio of cases to controls was 1:4 or greater, the greedy-descending and optimal matching strategies performed similarly. However, the optimal matching strategy took approximately 25 seconds to complete, while the greedy-descending strategy took approximately 0.9 seconds.

Ratio of cases to controls: As expected, increasing the ratio of cases to controls improved the proportion of matched cases. For the optimal, greedy-descending, and greedy-random strategies, 1:4 and 1:5 ratios of cases to controls yielded similar results, indicating diminishing returns as the number of controls relative to the number of cases becomes excessively large.

Number of variables in the model: The results illustrate that as the number of variables increases, the proportion of matched cases generally decreases. However, as the ratio of cases to controls increases, adding additional variables to the model has less of an effect on the proportion of matches.

Table 2: PROC PSMATCH results reported as mean % of cases matched (95% confidence interval)

		Number of Variables				
Cases:Controls (input data)	**Matching strategy**	**5**	**10**	**15**	**20**	**25**
1:1	Optimal	n/a	n/a	n/a	n/a	n/a
	Greedy					
	Ascending	60.5 (58.2, 62.8)	59.2 (56.9, 61.5)	58.7 (56.2, 61.0)	58.4 (55.9, 60.8)	58.5 (56.0, 60.8)
	Descending	70.0 (67.6, 72.6)	68.7 (66.3, 71.2)	68.0 (65.5, 70.7)	67.9 (65.4, 70.5)	67.9 (65.4, 70.4)
	Random	62.5 (60.2, 64.9)	61.0 (58.7, 63.2)	60.4 (58.1, 62.7)	60.2 (57.9, 62.4)	60.1 (57.9, 62.5)
1:2	Optimal	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	82.9 (80.6, 85.3)	82 (79.5, 84.4)	81.6 (78.9, 84.2)	81.3 (78.7, 83.7)	81.0 (78.5, 83.5)
	Descending	91.7 (89.6, 93.7)	90.4 (88.3, 92.5)	90.8 (88.6, 92.8)	90.4 (88.2, 92.5)	90.2 (88.0, 92.3)
	Random	84.6 (82.4, 86.7)	83.5 (81.4, 85.7)	83.3 (80.9, 85.5)	83.0 (80.8, 85.1)	82.7 (80.5, 85.0)
1:3	Optimal	100 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)
	Greedy					
	Ascending	91.5 (89.4, 93.5)	90.9 (88.6, 93.0)	91.0 (88.7, 93.0)	90.9 (88.6, 93.1)	90.8 (88.5, 92.9)
	Descending	97.9 (96.2, 99.5)	97.6 (95.8, 99.1)	97.4 (95.7, 99.0)	97.5 (95.7, 98.9)	97.4 (95.8, 98.9)
	Random	92.9 (90.9, 94.7)	92.2 (90.3, 94.1)	92.2 (90.3, 94.0)	92.1 (90.1, 94.1)	92.0 (90.0, 93.9)
1:4	Optimal	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.8, 100)	99.9 (99.7, 100)
	Greedy					
	Ascending	96.8 (94.9, 98.5)	96.1 (94.2, 97.8)	95.8 (93.9, 97.7)	96.0 (94.0, 97.7)	95.9 (94.0, 97.7)
	Descending	99.6 (99.1, 99.9)	99.5 (98.6, 99.9)	99.4 (98.5, 99.9)	99.5 (98.5, 99.9)	99.4 (98.4, 99.9)
	Random	97.6 (95.9, 99.0)	96.9 (95.3, 98.3)	96.7 (95.0, 98.3)	96.8 (95.1, 98.3)	96.7 (95.0, 98.2)
1:5	Optimal	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)	100 (99.9, 100)
	Greedy					
	Ascending	99.1 (97.7, 99.9)	98.6 (97.2, 99.7)	98.7 (97.2, 99.8)	98.6 (97.2, 99.8)	98.6 (96.9, 99.8)
	Descending	99.9 (99.6, 100)	99.8 (99.4, 100)	99.8 (99.5, 100)	99.8 (99.5, 100)	99.9 (99.5, 100)
	Random	99.3 (98.3, 99.9)	99.0 (97.9, 99.8)	98.9 (97.8, 99.7)	98.9 (97.7, 99.8)	98.8 (97.6, 99.7)

Note: "n/a" indicates that the model did not converge and the following error message was produced: "ERROR: A feasible optimal fixed ratio matching that has the specified parameters does not exist."

It is important to consider that, in addition to the number of variables in the model, *which* variables are included can also play a role in the proportion and quality of matches. Analyzing the effect of variable selection on the outcomes of propensity score matching was beyond the scope of this paper, but should be carefully considered.

CONCLUSION

This example illustrates the effects of the matching strategy, ratio of cases to controls in the input data set, and the number of variables in the model on the proportion of cases that are successfully matched. Overall, results revealed that increasing the number of variables in the model and decreasing the ratio of cases to controls in the input dataset results in fewer matched cases. The optimal and greedy-descending strategies produced similar proportions of matches in most circumstances, but the optimal strategy was substantially more computationally intensive.

REFERENCES

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. Statistics in medicine, 33(6), 1057-1069.

Yuan, Y., Yung, Y. F., & Stokes, M. Propensity Score Methods for Causal Inference with the PSMATCH Procedure.

SUPPLEMENTAL MATERIAL

The SAS Macro below provides an example for performing propensity score matching using each matching strategy and the different case to control ratios. This example shows the use of 5 variables but was cursively used for 10, 15, 20, and 25 variables.

```
libname c 'D:\File Location';

%let data = oneone onetwo onethree onefour onefive;

options mprint;
%macro matchymatchy(iteration);
%let i = 1;
%do %until(not %length(%scan(&data. ,&i)));
      proc psmatch data=%scan(&data., &i)_c region=allobs;
                  where Replicate = &iteration.;
            class group &Vars_Exact.;
            psmodel group(Treated="Case") = &Vars. &Vars_Exact.;
            match method=greedy(k=1 order = ascending) stat=lps;
            assess lps var=(&Vars.)
                  / weight=none plots=none;
            output out(obs=match)=&Vars_num._var_asc_&i. lps=_lps
matchid=_MatchID;
      run;
      proc psmatch data=%scan(&data., &i)_c region=allobs;
                  where Replicate = &iteration.;
            class group &Vars_Exact.;
            psmodel group(Treated="Case") = &Vars. &Vars_Exact.;
            match method=greedy(k=1 order = descending) stat=lps;
            assess lps var=(&Vars.)
                  / weight=none plots=none;
            output out(obs=match)=&Vars_num._var_des_&i. lps=_lps
matchid=_MatchID;
      run;
        proc psmatch data=%scan(&data., &i)_c region=allobs;
                  where Replicate = &iteration.;
            class group &Vars_Exact.;
            psmodel group(Treated="Case") = &Vars. &Vars_Exact.;
            match method=greedy(k=1 order= random(seed = 12345)) stat=lps;
            assess lps var=(&Vars.)
                  / weight=none plots=none;
            output out(obs=match)=&Vars_num._var_ran_&i. lps=_lps
matchid=_MatchID;
      run;
      proc psmatch data=%scan(&data. ,&i)_c region=cs;
                  where Replicate = &iteration.;
            class group &Vars_Exact.;
            psmodel group(Treated="Case") = &Vars. &Vars_Exact.;
            match method=optimal(k=1) exact = (&Vars_Exact.) distance=mah(lps
var=(&Vars.))caliper=.;
            assess lps var=(&Vars.)
                  / weight=none plots=none;
            output out(obs=match)=&Vars_num._var_opt_&i. lps=_lps
matchid=_MatchID;
      run;
  %let i = %eval(&i + 1);
```

```
  %end;
  %mend;

%Macro TablesOutput;
%let i = 1;
%do %until(not %length(%scan(&data. ,&i)));
      %let ii = 1;
      %do %until(not %length(%scan(&Methods. ,&ii)));
            proc means data = &Vars_num._var_%scan(&Methods. ,&ii)_&i.
noprint;
                  where group = "Case";
                  var male;
                  output out = c&Vars_num._var_%scan(&Methods. ,&ii)_&i. N =
c&Vars_num._var_%scan(&Methods. ,&ii)_&i.;
            run;
      %let ii = %eval(&ii + 1);
      %end;
%let i = %eval(&i + 1);
%end;
%Mend;

%Macro BS(num_group, its);
%let i = 1;
%do %until(not %length(%scan(&data. ,&i)));
      %let num_groupi = %eval(&num_group. * &i.);
      proc surveyselect data = %scan(&data. ,&i) out = %scan(&data. ,&i)_c n
= (&num_group. &num_groupi.) method = srs reps = &its. seed = 5101988;
            strata group;
      run;
%let i = %eval(&i + 1);
%end;
%do it = 1 %to &its.;
      %matchymatchy(&it.);
      %TablesOutput;
      data Summary_Data_;
            merge       c&Vars_num._var_asc_1-c&Vars_num._var_asc_5
                        c&Vars_num._var_des_1-c&Vars_num._var_des_5
                        c&Vars_num._var_ran_1-c&Vars_num._var_ran_5
                        c&Vars_num._var_opt_1-c&Vars_num._var_opt_5;
            it = &it.;
      run;
      data Summary_Data;
            set Summary_Data Summary_Data_;
      run;
      /* Clean log and output */
      dm log "clear";
      dm output "clear";
%end;
data Summary_Data_&Vars_num.;
      set Summary_Data;
      keep it c&Vars_num._var_asc_1-c&Vars_num._var_asc_5
                  c&Vars_num._var_des_1-c&Vars_num._var_des_5
                  c&Vars_num._var_ran_1-c&Vars_num._var_ran_5
                  c&Vars_num._var_opt_1-c&Vars_num._var_opt_5;
run;
%Mend;
```

```
%let Methods = asc des ran opt;

****************************************************************************;
****************************************************************************;
*FIVE VARIABLES;
****************************************************************************;
****************************************************************************;

%let Vars_num = Five;
%let Vars_Exact = male race;
%let Vars = age los CharlsScore;

data Summary_Data;
run;
/* Start timer */
%let _timer_start = %sysfunc(datetime());
ods exclude all;
options nomprint nosource nosource2 nonotes;
%BS(num_group = 1500, its = 1000);
ods exclude none;
options mprint source source2 notes;
/* Stop timer */
data _null_;
  dur = datetime() - &_timer_start;
  put 30*'-' / ' TOTAL DURATION:' dur time13.2 / 30*'-';
run;

data c.bs_summary_five_15Feb2019;
      set summary_data_five;
run;
```