

Evaluating the Utility of Coarsened Exact Matching for Pharmacoepidemiology Using Real and Simulated Claims Data

John E. Ripollone, Krista F. Huybrechts, Kenneth J. Rothman, Ryan E. Ferguson, and Jessica M. Franklin

WEB MATERIAL

Table of Contents

WEB TABLE 1. Distributions of the Non High Dimensional Propensity Score Covariates in the Pre-matched NSAID Cohort (Pharmaceutical Assistance Contract for the Elderly [United States, 1999-2002]) Dataset (n = 49,653).....	2
WEB TABLE 2. Distributions (Percentage) of the Covariates in the Pre-matched Statin Cohort (Medicaid Analytic eXtract [United States, 2000-2007]) Dataset (n = 886,996)	4
WEB TABLE 3. Summary of Plasmode Simulation Scenarios	6
WEB FIGURE 1. Plasmode Analysis Results, Non Interaction Scenarios – Variance, Including Coarsened Exact Matching Results.....	7
WEB FIGURE 2. Plasmode Analysis Results, Non Interaction Scenarios – Variance, Excluding Coarsened Exact Matching Results.....	8
WEB FIGURE 3. Plasmode Analysis Results, Non Interaction Scenarios – Bias, Including Coarsened Exact Matching Results.	9
WEB FIGURE 4. Plasmode Analysis Results, Non Interaction Scenarios – Bias, Excluding Coarsened Exact Matching Results.	10

WEB TABLE 1. Distributions of the Non High Dimensional Propensity Score Covariates in the Pre-matched NSAID Cohort (Pharmaceutical Assistance Contract for the Elderly [United States, 1999-2002]) Dataset (n = 49,653)

Covariate	Non-selective NSAIDs (n = 17,611)		Selective COX-2 Inhibitors (n = 32,042)	
	Mean (SD)	%	Mean (SD)	%
Age ^a	77.79 (7.30)		79.76 (7.24)	
Generics	7.43 (5.02)		8.41 (5.25)	
Any Medical Visit	7.74 (6.61)		8.60 (6.67)	
Charlson Comorbidity Score ^a	1.85 (1.97)		2.05 (2.01)	
Male ^a		18.84		14.09
Race				
White		89.76		95.45
Black		8.97		3.54
Other		1.27		1.02
Comorbidities				
Bleeding ^a		1.11		1.72
CHF		24.58		30.36
Coronary Disease		14.78		16.43
Hypertension		70.20		72.82
Rheumatoid Arthritis		2.70		5.00
Osteoarthritis		33.49		48.53
Ulcer ^a		2.42		3.71
Hospitalization in Prior Year		26.07		30.60
Nursing Home Resident		5.66		8.34
Other Medications				
Corticosteroid ^a		7.80		8.74
Other Gastrointestinal Medication ^a		20.44		27.42
Warfarin ^a		6.55		13.27
Year of Exposure Initiation				
1999		48.79		41.68
2000		23.91		29.94

2001	20.00	21.28
2002	7.30	7.09

Abbreviations: COX-2 = cyclooxygenase-2; NSAID = nonsteroidal anti-inflammatory drug.

^a This covariate appeared in the “very small” plasmode scenario

WEB TABLE 2. Distributions (Percentage) of the Covariates in the Pre-matched Statin Cohort (Medicaid Analytic eXtract [United States, 2000-2007]) Dataset (n = 886,996)

Covariate	Statins (n = 1,152)	No Statins (n = 885,844)
Age Categories		
≤ 19	5.56	29.43
20-24	14.06	35.6
25-29	21.09	20.41
30-34	28.13	9.48
35-39	22.22	4.17
≥ 40	8.94	0.91
Race		
Asian/Other Pacific Islander	6.51	3.42
Black/African American	25.69	34.09
Hispanic/Latino	17.10	15.08
Other	5.73	4.74
Unknown	2.95	2.01
White	42.01	40.67
U.S. Region		
Midwest	23.18	32.02
Northeast	21.27	14.97
South	26.04	26.07
West	29.51	26.94
Number of Non-antihypertensive Generics Used		
None	8.33	46.45
1-3	27.00	36.64
> 3	64.67	16.91
Number of Physician Visits During the Pre-index Period		
None	27.08	52.07
1-3	49.91	39.52
> 3	23.00	8.41

Year of Delivery		
2000	0.00	0.14
2001	4.17	9.65
2002	5.56	11.04
2003	10.42	14.59
2004	19.10	17.61
2005	20.14	16.88
2006	23.78	17.49
2007	16.84	12.60
Comorbidities		
Hypertension	40.63	5.00
Diabetes	45.14	3.06
Renal Disease	4.17	0.46
Obesity	23.35	5.31
Tobacco Use	11.02	7.77
Alcohol Abuse	3.99	2.61
Illicit Drug Use	6.42	5.33
Dyslipidemia	67.10	3.14
Multiple Gestation	6.60	3.55
Multipara	88.80	75.69
Other Medications		
Insulin	30.47	1.24
Antidiabetic	38.80	1.27
Hypertension Medication	53.73	6.65
Potentially Teratogenic Medication	31.68	3.63

WEB TABLE 3. Summary of Plasmode Simulation Scenarios

Exposure Prevalence ^a	Covariate Set ^b	Product Term ^c Estimate Strength
0.05	Very small Small Standard Large	
0.10	Very small Small Standard Large	
0.20	Very small Small Standard Large	Default Exaggerated
0.30	Very small Small Standard Large	
0.40	Very small Small Standard Large	

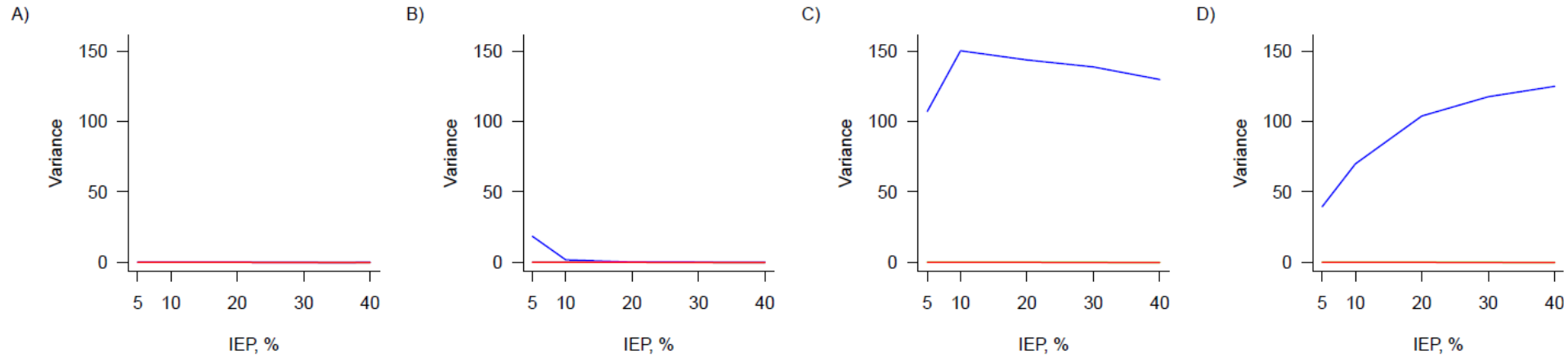
^a All plasmode scenarios were based on the NSAID cohort.

^b The “Very small” set comprised the 8 pre-determined covariates that were expected to be most associated with gastrointestinal complications; the “Small” set comprised all 19 pre-determined covariates; the “Standard” and “Large” sets comprised an additional 50 and 100 covariates, respectively, selected from a high-dimensional propensity score algorithm.

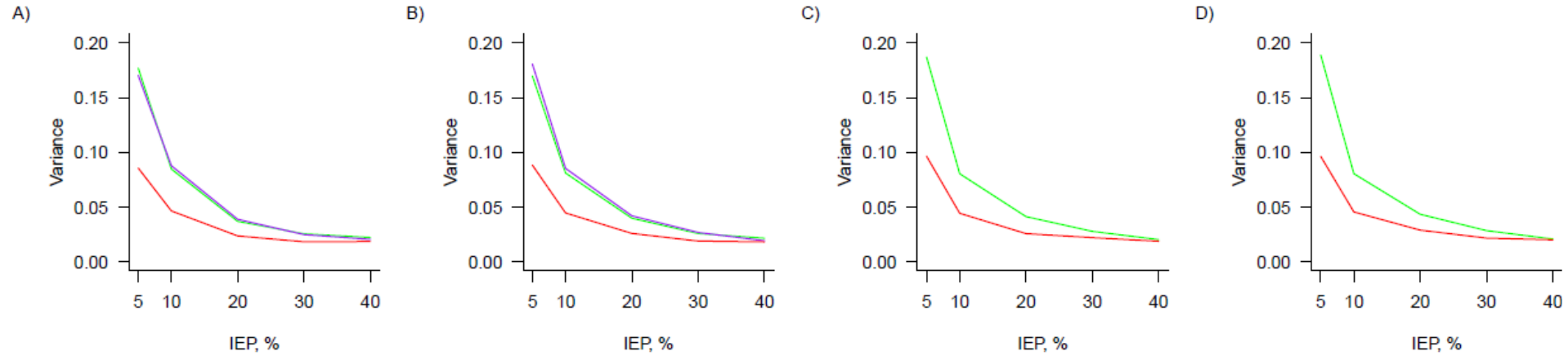
^c The product term represented the interaction between age and Charlson comorbidity score. The “default” scenario maintained the original product term and the “exaggerated” scenario was based on a product term that was 200% greater than the default product term.

WEB FIGURE 1. Plasmode Analysis Results, Non Interaction Scenarios – Variance, Including Coarsened Exact Matching Results.

A) Very small covariate set scenarios. B) Small covariate set scenarios. C) Standard covariate set scenarios. D) Large covariate set scenarios. Blue lines indicate coarsened exact matching trends, green lines indicate propensity score matching trends, purple lines indicate Mahalanobis distance matching trends and red lines indicate fine stratification by propensity score trends. IEP = Index exposure prevalence.

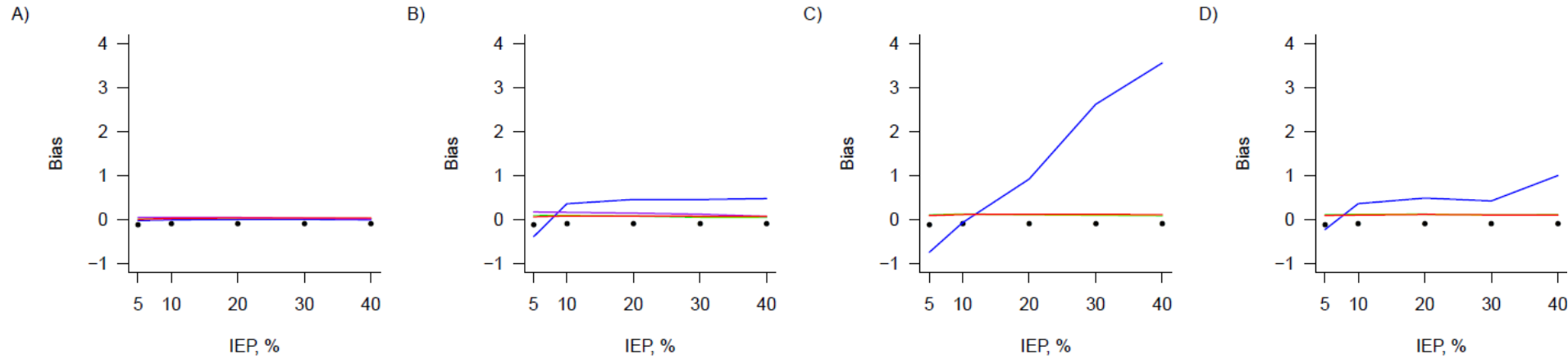


WEB FIGURE 2. Plasmode Analysis Results, Non Interaction Scenarios – Variance, Excluding Coarsened Exact Matching Results. A) Very small covariate set scenarios. B) Small covariate set scenarios. C) Standard covariate set scenarios. D) Large covariate set scenarios. Green lines indicate propensity score matching trends, purple lines indicate Mahalanobis distance matching trends and red lines indicate fine stratification by propensity score trends. IEP = Index exposure prevalence.



WEB FIGURE 3. Plasmode Analysis Results, Non Interaction Scenarios – Bias, Including Coarsened Exact Matching Results.

A) Very small covariate set scenarios. B) Small covariate set scenarios. C) Standard covariate set scenarios. D) Large covariate set scenarios. Black dots indicate the bias corresponding to the crude log risk ratio. Blue lines indicate coarsened exact matching trends, green lines indicate propensity score matching trends, purple lines indicate Mahalanobis distance matching trends and red lines indicate fine stratification by propensity score trends. IEP = Index exposure prevalence.



WEB FIGURE 4. Plasmode Analysis Results, Non Interaction Scenarios – Bias, Excluding Coarsened Exact Matching Results.

A) Very small covariate set scenarios. B) Small covariate set scenarios. C) Standard covariate set scenarios. D) Large covariate set scenarios. Black dots indicate the bias corresponding to the crude log risk ratio. Green lines indicate propensity score matching trends, purple lines indicate Mahalanobis distance matching trends and red lines indicate fine stratification by propensity score trends. IEP = Index exposure prevalence.

