INFORMS 2016

MUTUAL INFORMATION MINIMIZATION FOR EVALUATING THE CAUSAL IMPACT OF HOMECARE SERVICES ON PATIENT DISCHARGE DISPOSITION

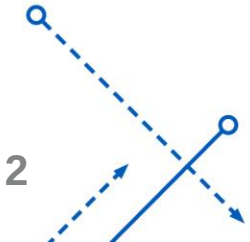
S. Casucci, L. Sun, Y. Zhou, A. Nikolaev, L. Lin

University at Buffalo
Department of Industrial and Systems Engineering
School of Engineering and Applied Sciences

Driving Forces

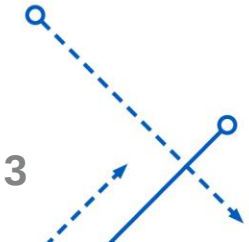
- Need for new care delivery models tailored for high risk patient populations
  - Reduce healthcare utilization
  - Improve health outcomes

- Adoption of a population health management approach
  - Allows for predictive models that consider multiple chronic conditions
    - Must consider the complex interactions between diseases, patients, treatment methods, and care delivery logistics



Vulnerable Populations

- Low-income Medicaid population has unique health care challenges
  - High rates of chronic disease
    - Prevalent conditions include: behavioral health issues, diabetes, and hypertension
    - 73% of WNY Medicaid population has one or more chronic diseases in 2013
  - High rates of inpatient hospital care and subsequent readmissions
    - In 2013, 83% of the WNY Medicaid population experiencing a hospital readmission within 30 days of their initial in-patient treatment had one or more chronic diseases

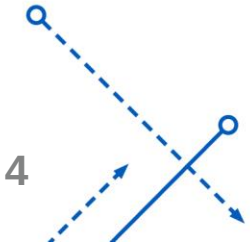


Reducing healthcare utilization and hospital readmissions

- Home health care
  - Includes a wide range of health care services provided in the home
  - Intended to treatment an illness or injury and help patients recover following a hospitalization[1]
  - Treatment in the home is typically less expensive and more convenient for individuals with chronic disease

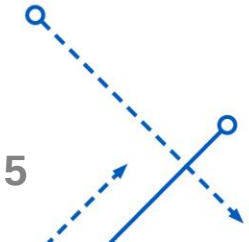
- Challenge
  - Services are typically designed for the elderly (65+) Medicare population

[1]https://www.medicare.gov/what-medicare-covers/home-health-care/home-health-care-what-is-it-what-to-expect.html



Study Objective

Develop health informatics based analytical models that will ultimately inform the design of home health care services for low-income adult Medicaid patients



Approach

Specific Aim #1 (Identifying prevalent chronic diseases and disease combinations)

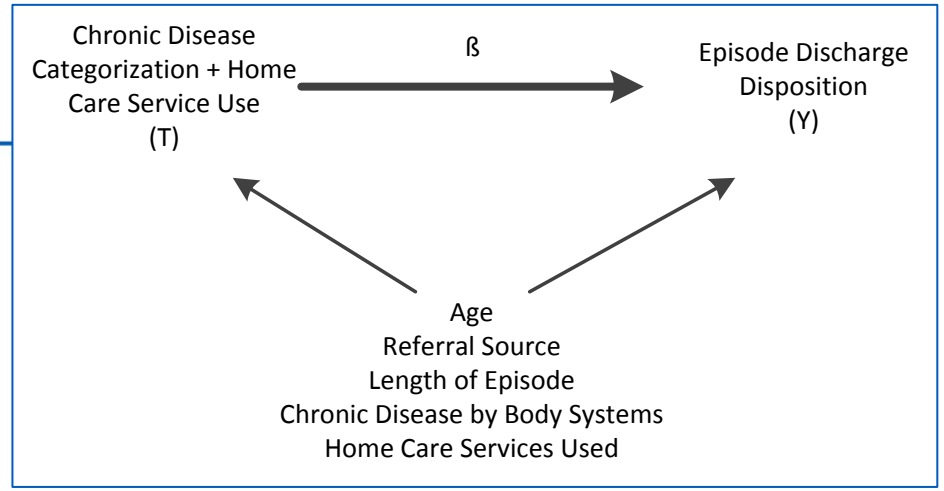
- Existing literature
- Descriptive statistics

Specific Aim #2 (Outcome evaluation for individuals with home health care services)

- Descriptive analysis of home care administrative and clinical data sets
- Inferential causal effect analysis

Specific Aim #3 (Opportunities to improve home health care services)

- Synthesis of information

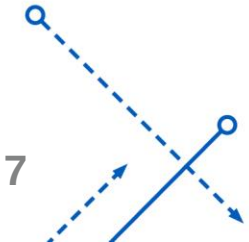


Causal Inference

- Measures the difference in potential outcomes at different levels of treatment

Rubin Causal Model

- Challenge
  - Only one potential outcome can be observed for a given individual (with or without treatment)
- Solution
  - Average Treatment Effect for the Treated (ATT)
    - Compare the averages across treatment and control groups for those who have actually received the treatment

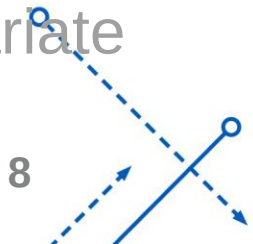


Matching

- Process of ensuring that covariate distributions in the treatment and control groups are a similar as possible
  - Treatment and control groups will differ only in their treatment indicators

Mutual Information (MI)

- A measure of the level of dependence between random variables
  - High values indicate high dependence – that one variable contains much information about the other
  - Low values indicate low dependence – that variables are independent
- In matching the difference between the covariate distributions can be directly measured by MI
  - Randomization in treatment assignment implies MI = 0 between covariate and treatment variables



Home Care Service Administrative Data

Unique episodes in data set (services received in 2014) 15,143

11,028 unique patients received home care services in 2014
65% of episodes followed a hospital discharge

Adults aged 18 or older 13,835

Study focuses on low-income adult Medicaid population (aged 18+)
28% of episodes had patient age < 65 years of age

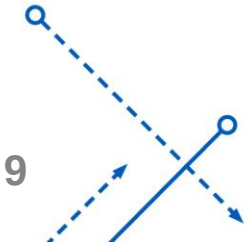
Start of Care Date on or after November 1, 2013

Average length of home care episode = 37 days
Standard deviation = 34 days
Range = 0 to 417 days

Episode discharge disposition is home or admitted to hospital 11,568
STUDY DATA SET

7% of episodes resulted in a hospital admission disposition
Percent of episodes with body systems affected by chronic disease

- One body system 12%
- Two body systems 52%
- Three body systems 36%



Identifying prevalent chronic diseases in study population

Healthcare Cost and Utilization Project (HCUP) Chronic Condition Indicator[1]

- Classifies all ICD-9-CM diagnosis codes
  - Chronic or non-chronic conditions
  - Body System affected

Body Systems Classifier
1 - Infectious and parasitic disease
2 - Neoplasms
3 - Endocrine, nutritional, and metabolic diseases and immunity disorders
4 - Diseases of blood and blood-forming organs
5 - Mental disorders
6 - Diseases of the nervous system and sense organs
7 - Disease of the circulatory system
8 - Diseases of the respiratory system
9 - Diseases of the digestive system
10 - Diseases of the genitourinary system
11 - Complications of pregnancy, childbirth, and the puerperium
12 - Diseases of the skin and subcutaneous tissue
13 - Diseases of the musculoskeletal system
14 - Congenital anomalies
15 - Certain conditions originating the perinatal period
16 - Symptoms, signs, and ill-defined conditions
17 - Injury and poisoning
18 - Factors influencing health status and contact with health services

[1]https://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp

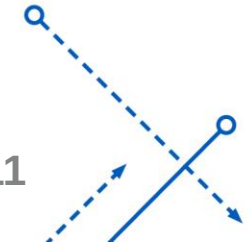
Identifying prevalent chronic diseases in study population

Existing Literature[1]

- Home health care patients (2000):
  - 7 – Diseases of the Circulatory System (26%)
  - 3 – Endocrine, nutritional, and metabolic diseases and immunity disorders (11%)
  - 13 – Diseases of the musculoskeletal system and connective tissue (10%)

Study Population

- 7 – Diseases of the Circulatory System (42%)
- 3 – Endocrine, nutritional, and metabolic diseases and immunity disorders (13%)
- 13 – Diseases of the musculoskeletal system and connective tissue (27%)



[1]https://www.cdc.gov/nchs/data/nhhcsd/curhomecare00.pdf

Experimental Design

Tier 1 Experiments

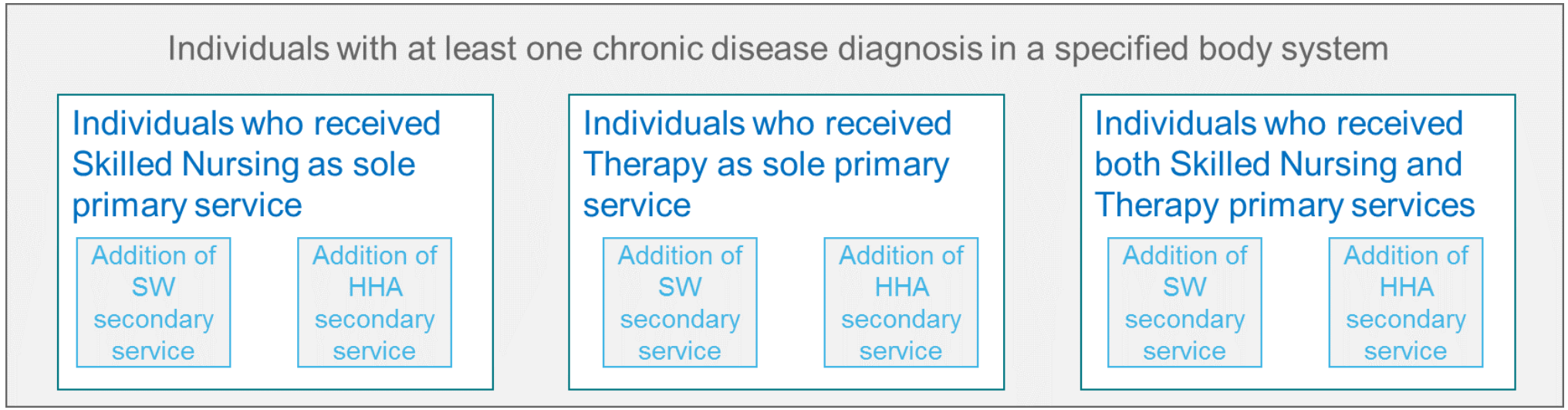
- Effect of primary (skilled) services on end of episode disposition

Tier 2 Experiments

- Effect of adding a secondary (unskilled) service on end of episode disposition

Primary (skilled) services
- Skilled nursing
- Therapies (physical, speech, occupational)

Secondary (unskilled) services
- Social work
- Home health aide (assistance with activities of daily living; eating, bathing, etc.)



MIM Method

- Mutual information based matching method (MIM)
- Non-parametric matching method
  - Frames matching as an optimization problem: select a subset $S \subseteq C$ that minimizes the MI between the treatment indicator and covariate vector over set $S \cup T$.

$$I(t;X) = \log(|\mathcal{S}| + |\mathcal{T}|) + \frac{1}{|\mathcal{S}| + |\mathcal{T}|}\Big(\sum_{b\in B} S_b[\log S_b - \log(T_b + S_b) - \log|\mathcal{S}|] + \sum_{b\in B} T_b[\log T_b - \log(T_b + S_b) - \log|\mathcal{T}|]\Big).$$

    - T = given treatment group
    - C = control pool
    - u = observed unit, $u \in T \cup C$ with covariate vectors $\{X_1, X_2, \ldots, X_{|K|}\}$

Optimality of MIM

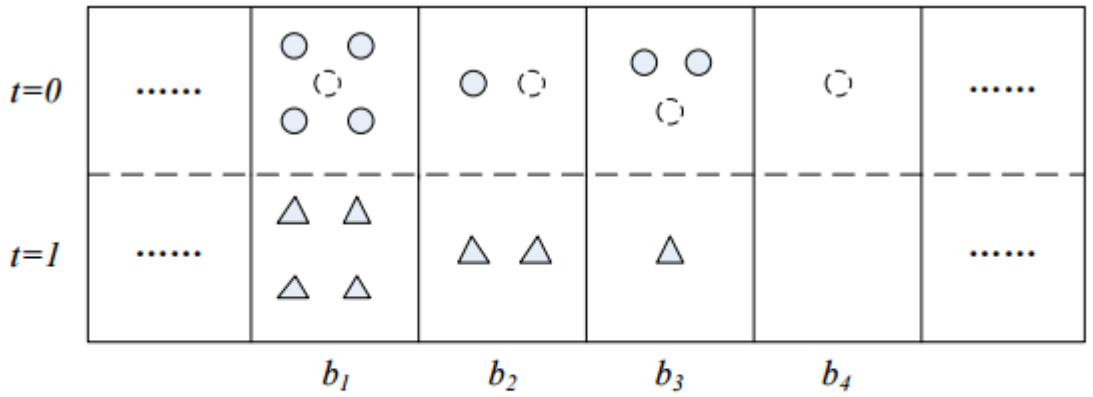


Figure 2: A selection process illustration. Treated units in $\mathcal{T}$, selected control units in $\mathcal{S}$ and unselected control pool units are represented by triangles, full circles and dashed circles, respectively.

- Theorem. Necessary and Sufficient Condition for Optimality
  - Consider an instance of $\min I(t;X)$. A control group S of size N is optimal if and only if for any pair of bins b_1 and b_2 with $|C_{b_2} - S_{b_2}| \geq 1$ it holds that $\frac{S_{b_1}-1-A}{T_{b_1}} \leq \frac{S_{b_2}-A}{T_{b_2}}$, where $A \approx -0.47$.

Sequential Selection Algorithm for MIM

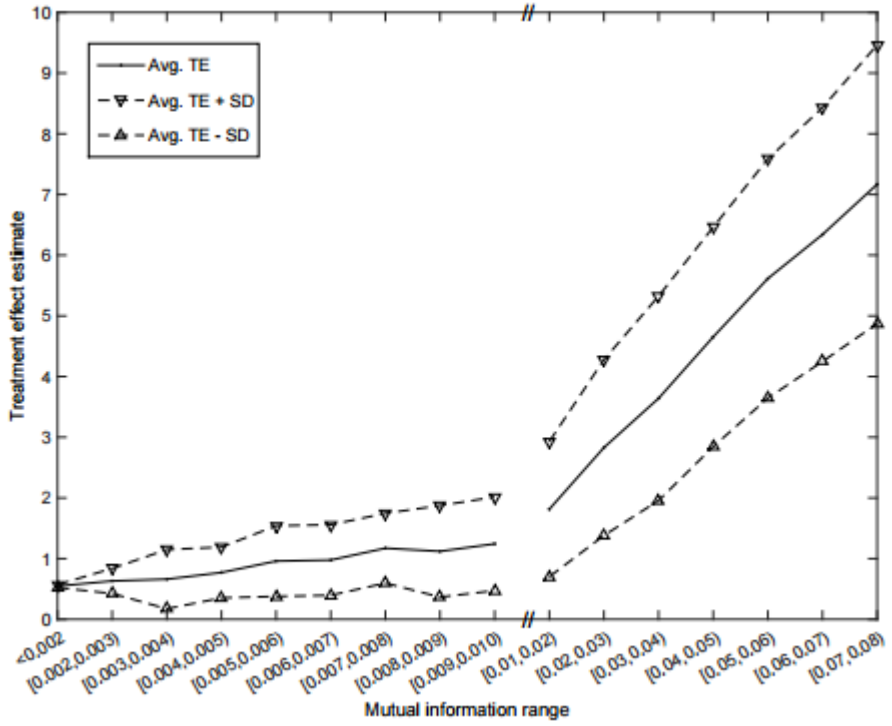
- Theorem
  - If control group S has the minimum $I(t;X)$ among all the control groups of size N, then a group with the minimum $I(t;X)$ among all the control groups of size $N+1$ can be obtained from S by adding to it a single unit from bin $b \in \operatorname{argmin}_{b\in B}\{\frac{S_b - A}{T_b} : |C_b - S_b| \geq 1\}$.
- Algorithm

1: Initialize the joint bin set $\{b : C_b + T_b \geq 1\}$ consisting of all the bins occupied by the units in $\mathcal{T} \cup \mathcal{C}$; compute T_b and C_b; set $S_b = 0$ for all b.
2: Select a bin $b \in \operatorname{argmin}_{b\in B}\{\frac{S_b - A}{T_b} : |C_b - S_b| \geq 1\}$, update S_b by adding 1.
3: If N units are selected, go to step 4. Otherwise, go to step 2.
4: Construct a control group complying with the obtained values of S_b over all the initialized bins b. Stop.

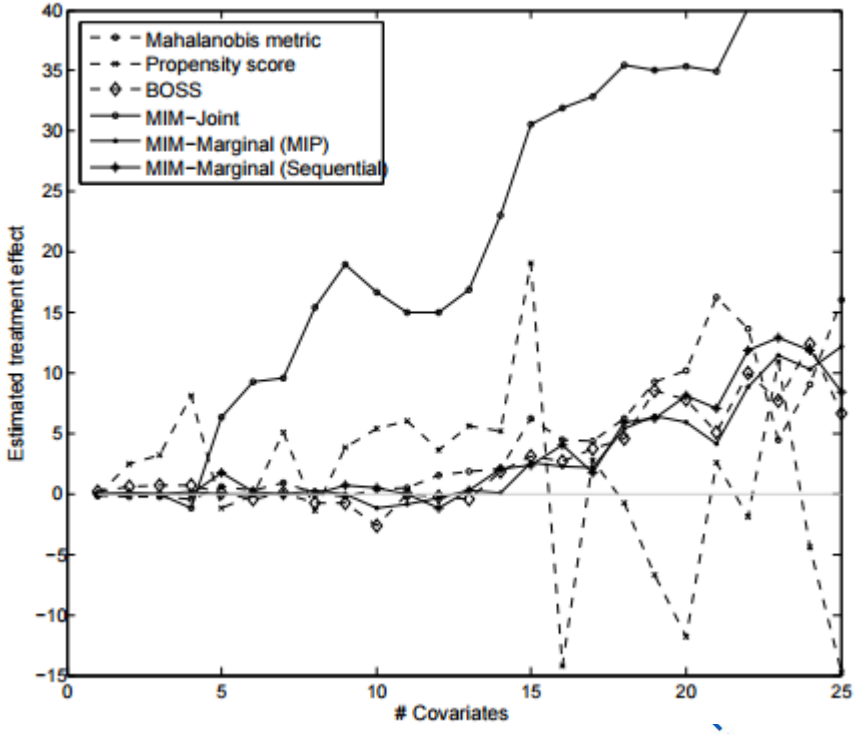
MIM Solution Quality

A synthetic data set with 25 covariates,100 treatment units and 10,000 control units is generated with all the covariate values drawn from normal distributions with mean 0. All the treatment and control units have the same, highly nonlinear response function.

ATT vs MI (10 covariates)

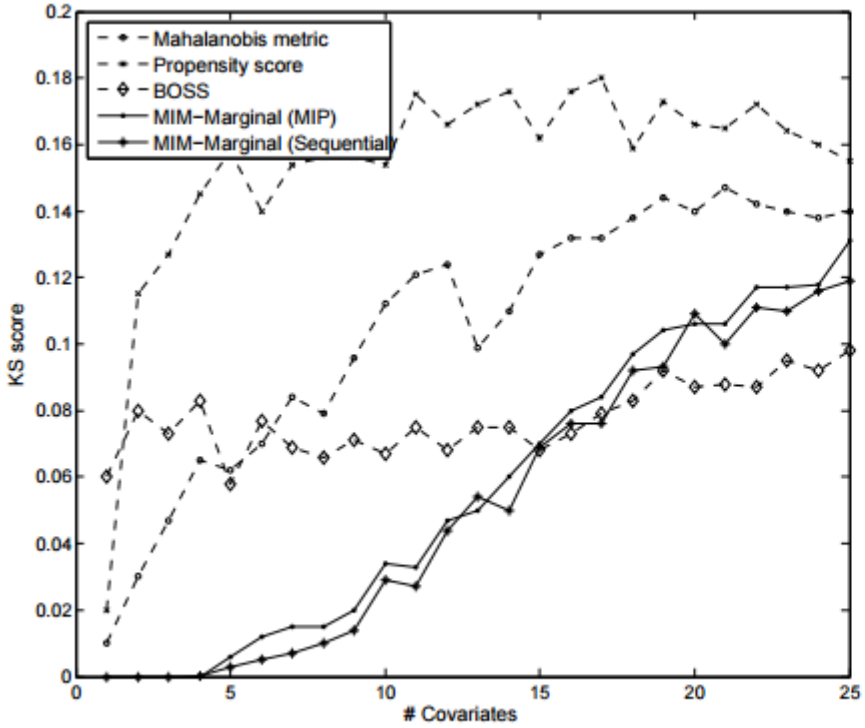


ATT estimates with different methods

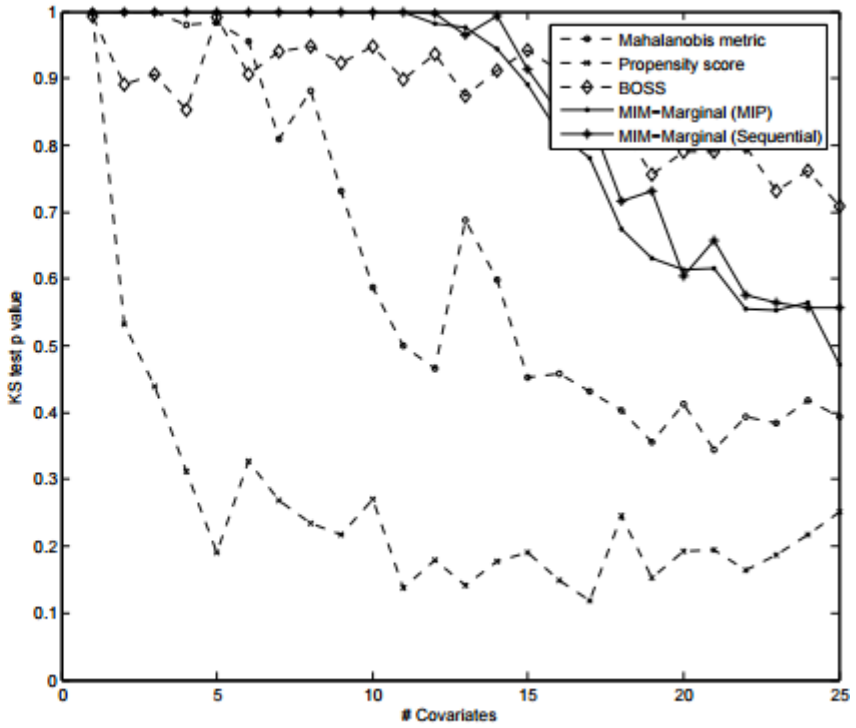


MIM Solution Quality

- KS test for marginal balance quality



(a) KS scores



(b) KS test p values

MIM Time Performance

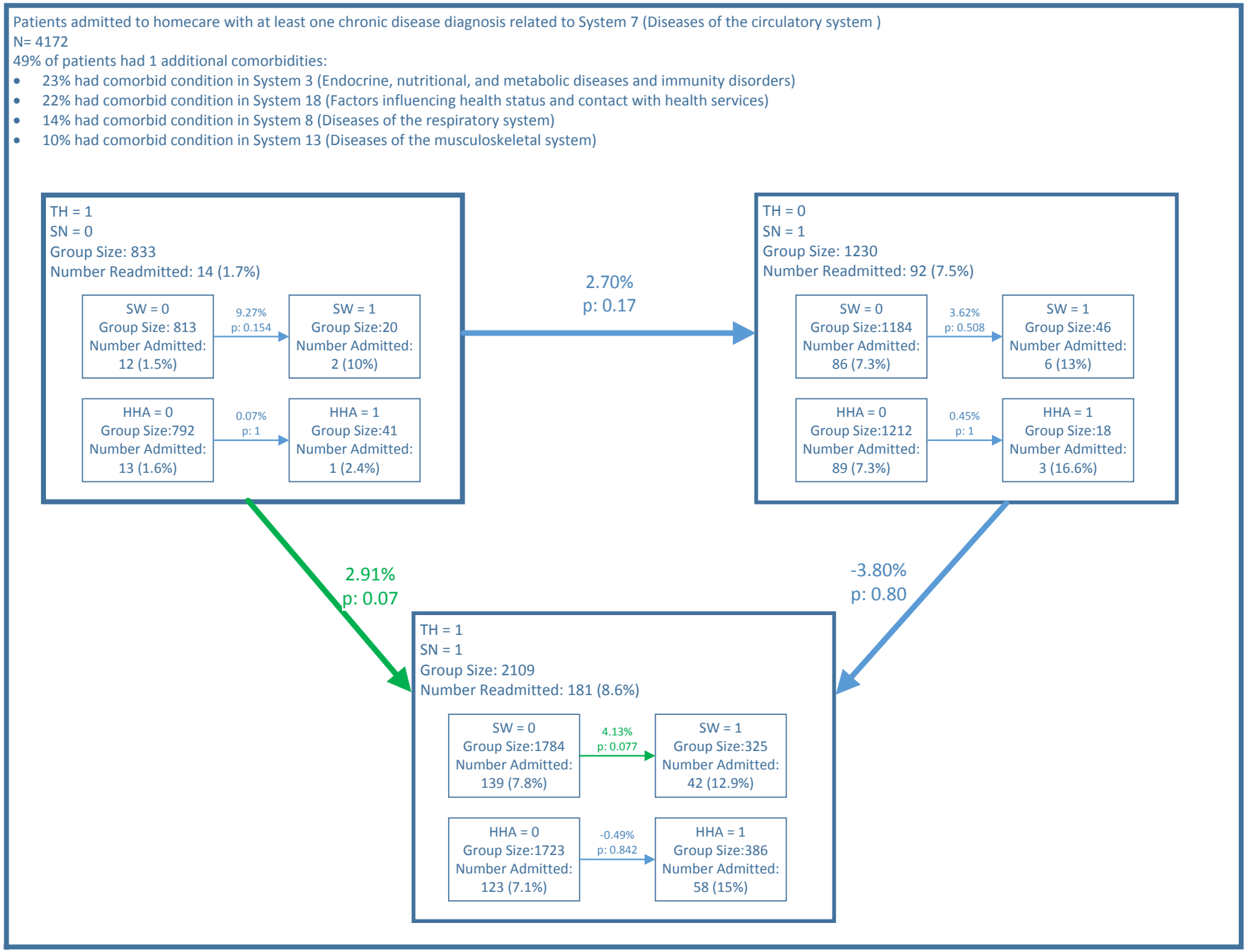
- Computational efficiency of the MIM-Marginal method is unparalleled, making it highly practical for large-scale data mining.

Table 3: Performance of the MIM-Marginal method on practical data sets.

		Census 94-95 (US Citizenship)	Census 94-95 (Business Ownership)	Extracted 94 (US Native)	Extracted 94 (Doctorate)	NSW 86 (Training)
Data	# Control Units	186,122	196,825	29,170	32,148	15,992
	# Treated Units	13,401	2,698	3,391	413	185
	# Continuous Covariates	8	8	6	6	4
	# Categorical Covariates	32	32	7	7	4
	# Marginal Bins	645	647	180	206	88
Balance	Avg KS	0.064	0.001	0.020	0.095	0.032
	Max KS	0.412	0.002	0.133	0.200	0.049
	Avg p-value	0.524	1.000	0.769	0.746	0.995
	Min p-value	0.109	1.000	0.518	0.459	0.981
	Avg MI	0.020	0.001	0.004	0.085	0.002
Time	Binning (seconds)	4,410.84	4,319.71	46.23	47.62	1.74
	Matching (seconds)	66,950.58	14,235.49	399.70	49.83	1.76

System 7 (Diseases of the Circulatory System)

- Care plans should include Therapy (TH) services and Skilled Nursing (SN) to reduce end of episode dispositions of ‘admit to hospital’
- The addition of social work (SW) services further will further decreases end of episode disposition of ‘admit to hospital’.

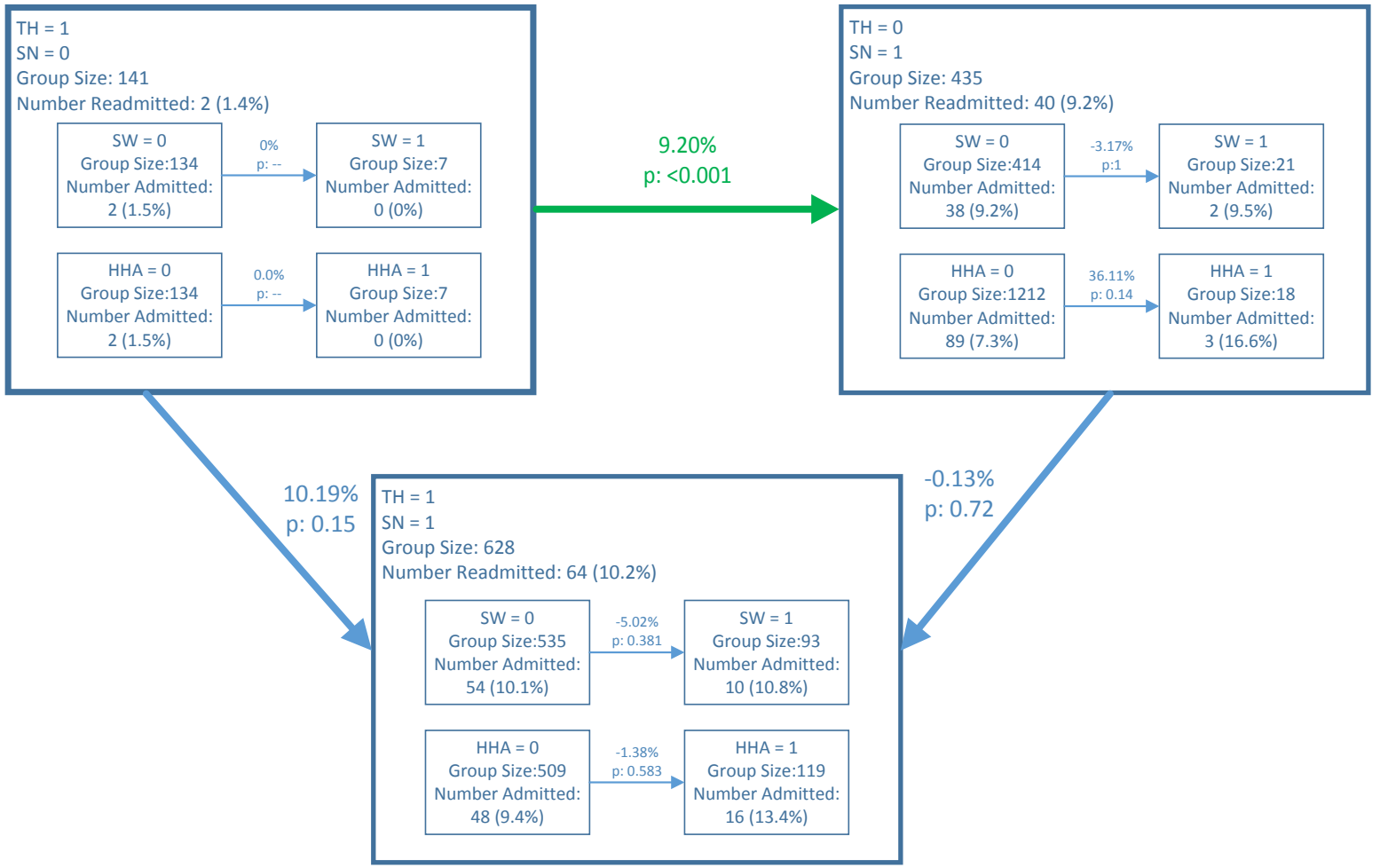


System 3 (Endocrine, nutritional, and metabolic diseases and immunity disorders)

- Care plans should include Skilled Nursing (SN) service only with no additional secondary services to reduce on average the rate of end of episode dispositions of ‘admit to hospital’.

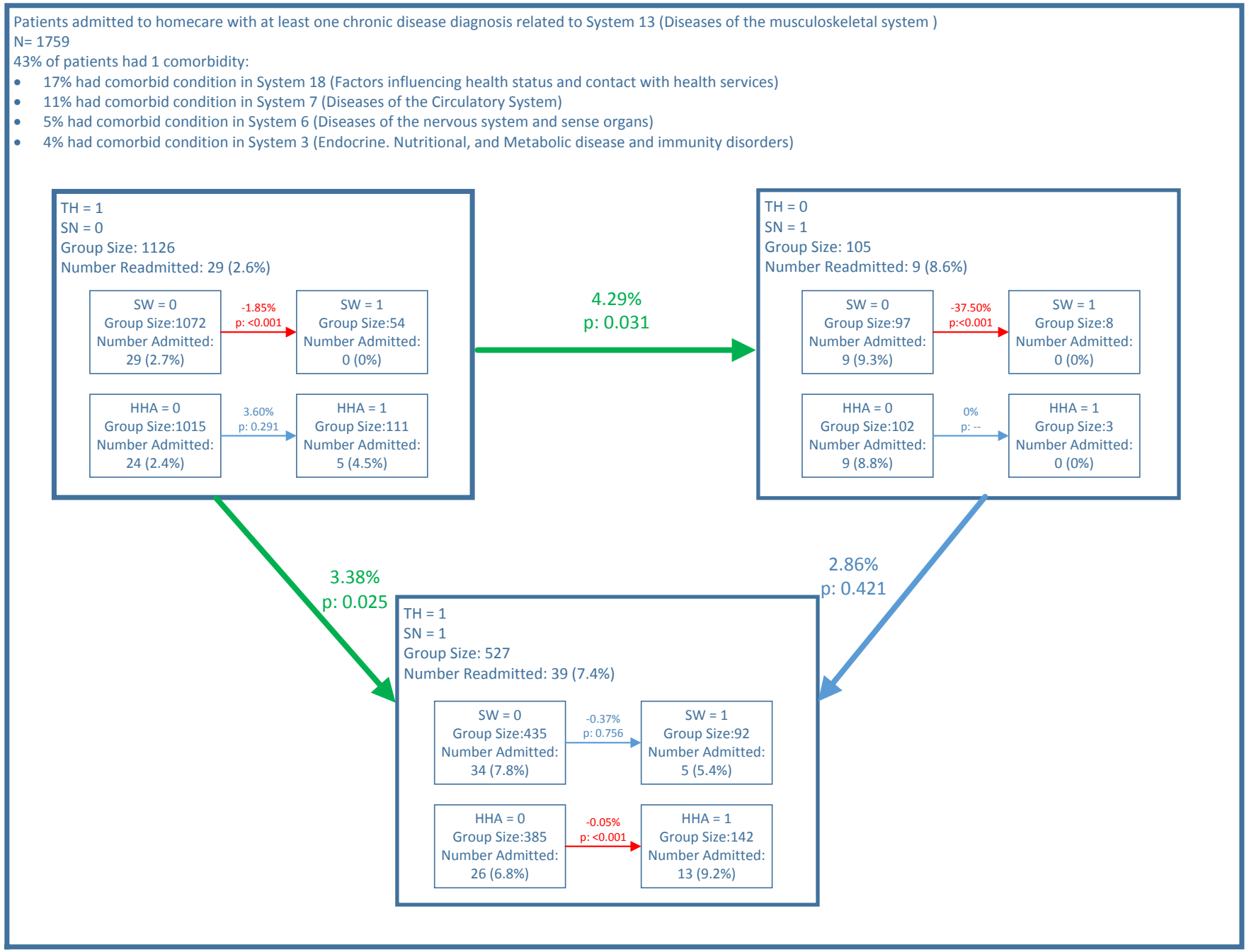
Patients admitted to homecare with at least one chronic disease diagnosis related to System 3 (Endocrine. Nutritional, and Metabolic disease and immunity disorders)
N= 1024
75% of patients had 1 comorbidity:

- 54% had comorbid condition in System 7 (Diseases of the Circulatory System)
- 8% had comorbid condition in System 18 (factors influencing health status and contact with health services)
- 8% had comorbid condition in System 12 (Diseases of the skin and subcutaneous tissue)
- 7% had comorbid condition in System 13 (Diseases of the musculoskeletal system)



System 13 (Diseases of the musculoskeletal system and connective tissue)

- Care plans should include Skilled Nursing (SN) service only or Skilled Nursing (SN) and Therapy (TH) to reduce on average the rate of end of episode dispositions of ‘admit to hospital’
- The addition of any secondary service has little effect or, in some cases, a negative effect (increase the rate of ‘admit to hospital’ dispositions and therefore should not be included in the care plan.

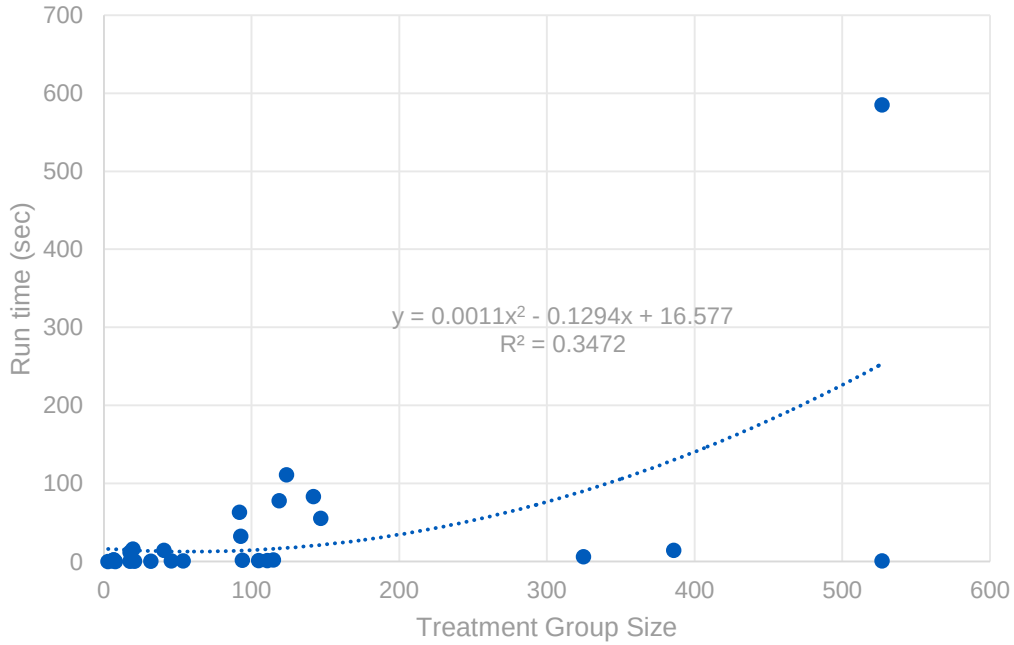


MIM in the Present Application

Early Trials

- Used iterative genetic matching algorithm[1]
  - Run times
    - Matching – polynomial increase
    - GenMatch step – exponential increase
  - Experimental run times
    - Ranged from 70 – 91 hours

- MIM Run times
  - Ranged from 0.02 – 585 seconds



[1]Sekhon JS. Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software.* 2011;42(7):1-52.

Conclusion

The newly developed information theory-driven causal inference method allowed for large-scale exploratory analyses of health care data, i.e., for testing multiple causal hypotheses in reasonable time to shed practical insights.

The findings about best practices of home care will help inform health care policy-making.

University at Buffalo
Department of Industrial and Systems Engineering
School of Engineering and Applied Sciences

THANK YOU

Questions?..