Evaluating the Clinical Phenotypes Associated with Progression to Chronic Kidney Disease Using an Unsupervised Machine Learning Approach

Mohammad Al-Mamun, PhD [1] Todd Brothers, Pharm D, BCCCP, BCPS [2]
Ki Jen Jeun [1] Imtiaz Ahmed [3]

[1] The University of West Virginia, School of Pharmacy [2] The University of Rhode Island, College of Pharmacy
3 The University of West Virginia, Department of Industrial and Management Systems Engineering

Introduction

The association between the initial occurrence of AKI and long-term CKD risk is multifactorial and complex.[1] Assessing the risk of AKI and its acceleration towards CKD and ESRD remains an area of research to evaluate the etiology, influential clinical biomarkers, severity, and recurrence.[2,3] However, no studies have comprehensively investigated the clinical phenotyping using traditional and non-traditional risk factors.

Purpose

Our primary aim of this retrospective observational cohort study was to examine non-traditional clinical phenotypes of Acute Kidney Injury (AKI) leading to Chronic Kidney Disease (CKD) using an unsupervised data mining and graphical network approach.

Methodology

90,602 West Virginia University Health System electronic patient records were examined from February 1, 2010, to June 30, 2022. Patients excluded were individuals that received renal replacement therapy prior to or had an AKI event less than 3 years prior to CKD diagnosis. AKI status was determined according to the 2012 Kidney Disease Improving Global Outcomes classification system. (Figure 1)

Figure 1: 2012 Kidney Disease Improving Global Outcomes (KDIGO) classification system.

Stage	SCr	UO
1	1.5-1.9 times baseline OR ≥ 0.3 mg/dL increase	<0.5mL/kg/hr for 6-12 hours
2	2.0-2.9 times baseline	<0.5mL/kg/hr ≥12hrs
3	3.0 times baseline OR Increase SCr to >4 mg/dL	<0.3mL/kg/hr for ≥24 hrs OR Anuria for ≥12 hrs

Hospital-Acquired AKI (HA-AKI) and Community-Acquired AKI (CA-AKI) were defined as an AKI event within or 90-days after hospitalization prior to CKD diagnosis, respectively. A hierarchical clustering method was used to identify, and a graphical network was used to confirm phenotype clusters, respectively. Phenotypic networks were constructed for each cohort with each node representing a disease diagnosis (comorbidity) and each edge denoting the co-occurrence relationship between pair-wise comorbidities. A centrality metric of 'betweenness' was used to identify key nodes between components within the network.

Results

Table 1 describes baseline characteristics related to HA- and CA-AKI. Individuals were predominantly White, female, and with mean age of 61 years. Non-AKI group had a higher number of comorbidities. Common AKI features were long-term opiate use, atelectasis, ischemic heart disease, lactic acidosis, diabetes.

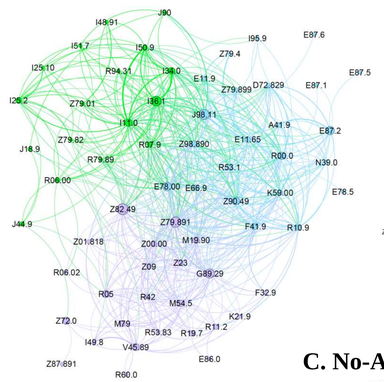
Table 1: Baseline Characteristics

	HA-AKI	CA-AKI	No-AKI	*p*-value [a]
	n = 5981	n = 6762	n = 46133	
Sex				<0.01
Male	2902 (48.5%)	3284 (48.6%)	24448 (53.0%)	
Female	3077 (51.4%)	3475 (51.4%)	21671 (47.0%)	
Unknown	2 (0.03%)	3 (0.04%)	14 (0.03%)	
Race				<0.01
White	217 (3.63%)	300 (4.44%)	1658 (3.59%)	
Non-White	442 (7.39%)	487 (7.20%)	6303 (13.7%)	
Unknown	5322 (89.0%)	5975 (88.4%)	38172 (82.7%)	
Ethnicity				<0.01
Hispanic/Latino	17 (0.28%)	23 (0.34%)	175 (0.38%)	
Non-Hispanic/Latino	5346 (89.4%)	6128 (90.6%)	37314 (80.9%)	
Unknown	618 (10.3%)	611 (9.04%)	8644 (18.7%)	
Age				<0.01
Mean	65.1 (13.6)	66.4 (13.2)	68.6 (13.4)	
Ages by group:				<0.01
< 44	460 (8.19%)	423 (6.68%)	2429 (5.81%)	
45-64	1994 (35.5%)	2063 (32.6%)	10707 (25.6%)	
> 64	3161 (56.3%)	3848 (60.8%)	28662 (68.6%)	
Comorbidity				
Anemia	1651 (27.6%)	2181 (32.3%)	4648 (10.1%)	<0.01
Atrial fibrillation	187 (3.13%)	178 (2.63%)	3651 (7.91%)	<0.01
Anxiety	981 (16.4%)	1244 (18.4%)	8703 (18.9%)	<0.01
COPD	936 (15.6%)	1192 (17.6%)	8203 (17.8%)	<0.01
Heart Failure	1192 (19.9%)	1500 (22.2%)	7425 (16.1%)	<0.01
Hyperlipidemia	1182 (19.8%)	1567 (23.2%)	17424 (37.8%)	<0.01
Hypertension	552 (9.23%)	972 (14.4%)	35184 (76.3%)	<0.01
Nicotine dependence	865 (14.5%)	1262 (18.7%)	15533 (33.7%)	<0.01
Obesity	899 (15.0%)	1308 (19.3%)	9595 (20.8%)	<0.01
T2DM	2625 (43.9%)	3146 (46.5%)	17919 (38.8%)	<0.01

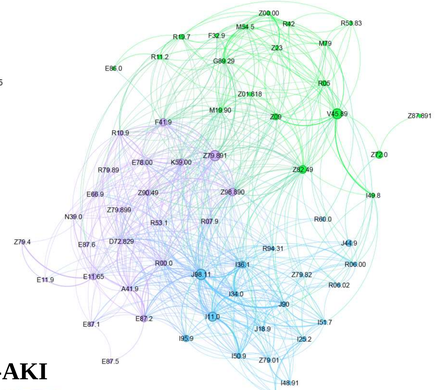
[a] Baseline characteristics of HA-AKI, CA-AKI, and No-AKI patients were compared using Pearson chi-square tests for categorical variables and independent-samples *t*-tests for continuous variables.

Figure 1 depicts the Gephi graphs of the network analytics for comorbidities and procedures in the three cohorts (HA-AKI, CA-AKI, and No-AKI).

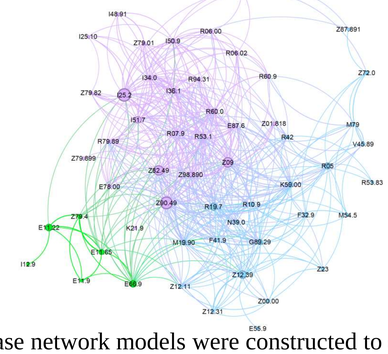
A. Hospital-Acquired AKI



B. Community-Acquired AKI



C. No-AKI



Phenotypic disease network models were constructed to evaluate risk factor profiles among the cohorts. The network analysis provides a graphical representation of the complex pattern relationships among the risk factors. For example, multiple nodes (ICD-9/10 codes) were identified in cohorts A and B, but not cohort C. Particularly, Z79.891 (long-term opiate analgesic use), J98.11 (atelectasis), Z82.49 (ischemic heart disease), E87.2(lactic acidosis) and E11.65 (Type 2 diabetes mellitus).

Conclusion

The findings of our study suggest that incorporation of data mining methodologies may lead to improved identification of otherwise unknown clinical phenotypes to reduce the transition of HA and CA-AKI to CKD.

References

1. James MT, Bhatt M, Pannu N, Tonelli M. Long-term outcomes of acute kidney injury and strategies for improved care. *Nat Rev Nephrol*. 2020;16(4):193–205.
2. Coca SG, Singanamala S, Parikh CR. Chronic kidney disease after acute kidney injury: a systematic review and meta-analysis. *Kidney Int*. 2012;81(5):442–448.
3. Siew ED, Parr SK, Abdel-Kader K, Eden SK, Peterson JF, Bansal N, Hung AM, et.al. Predictors of recurrent AKI. *J Am Soc Nephrol*. 2016;27: 1190–1200.

The authors have no conflicts of interest concerning this research