Application of group LASSO regression based Bayesian networks in risk factors exploration and disease prediction for acute kidney injury in hospitalized patients with hematologic malignancies

Yang Li

Zhongshan Hospital Fudan University

Xiaohong Chen

Zhongshan Hospital Fudan University

Yimei Wang

Zhongshan Hospital Fudan University

Jiachang Hu

Zhongshan Hospital Fudan University

Ziyan Shen

Zhongshan Hospital Fudan University

Xiaoqiang Ding (✉ dingxqzshospital@163.com)

Zhongshan Hospital Fudan University

Research article

Keywords: Acute kidney injury, Hematologic malignancy, Bayesian network, Disease prediction, Clinical epidemiology

Posted Date: December 3rd, 2019

DOI: <https://doi.org/10.21203/rs.2.16338/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Nephrology on May 5th, 2020. See the published version at <https://doi.org/10.1186/s12882-020-01786-w>.

Abstract

Background: This study aims to explore a novel machine-learning algorithm, Bayesian networks (BNs), to delineate the interrelationships between acute kidney injury (AKI) and its associated risk factors among patients with hematologic malignancies (HM), to assess the prediction ability of BNs model, and to infer the probability of AKI under different clinical settings. **Methods:** From 1 October 2014 to 30 September 2015, 2501 hospitalized patients diagnosed with HM in Zhongshan Hospital, Fudan University, Shanghai of China, were recruited in this retrospective study. Data on demographics, comorbidities, and baseline clinical lab records were exported from the electronic medical records. The group-LASSO (gLASSO) regression was performed to select the candidate predictors of AKI, which were further presented in BNs analysis for interrelationship exploration and disease prediction. **Results:** Among 2395 eligible patients, 370 episodes were diagnosed with AKI (15.4%). Patients with multiple myeloma (24.1%) and leukemia (23.9%) shared a higher AKI incidence than lymphoma (13.4%). Screened by the gLASSO regression, variables as age, gender, diabetes, hemopathy category, anti-tumor treatment, hemoglobin, serum creatinine (SCr), estimated glomerular filtration rate (eGFR), serum uric acid, serum sodium and potassium level were found with significant associations with AKI occurrence. BNs model revealed a complex interrelationship between these factors, among which there were direct connections between AKI and age, hemoglobin, eGFR, serum sodium and potassium. Moreover, HM category and anti-tumor treatment were indirectly linked to AKI through hemoglobin and eGFR; diabetes was connected with AKI through serum sodium level. BNs inferences indicated that when poor GFR, anemia and hyponatremia occurred simultaneously, the probability of AKI might reach 78.5%. The area under the receiver operating characteristic curve (AUC) of BNs model was 0.835, higher than that in logistic score model (0.763) and showed a robust performance in 10-fold cross-validation. **Conclusion:** AKI is common in HM hospitalized patients and is affected by multiple risk factors. The application of Bayesian networks and gLASSO regression can provide a novel strategy to explore the potential risk factors integrally and deep into their interrelationships, in accordance with the tide of e-alert and big-data for AKI early detection.

Background

Acute kidney injury (AKI) is one of the common complications in patients with hematologic malignancies (HM). A population-based cohort study in Danish revealed that 1-year risk of AKI was estimated as 18.8% in patients diagnosed with lymphoma, 27.5% in leukemia and 31.8% in multiple myeloma [1]. In these HM patients, the pathogenesis of AKI is usually multifactorial, partially similar with those of non-cancer patients, and also related to the underlying malignancies and subsequent treatment [2, 3]. The progression of AKI further limits anti-tumor treatment and brings about a higher in-hospital mortality and heavier economic burden [4, 5]. Because of the indifference to AKI prevention and the lack of awareness for repeated serum creatinine (SCr) testing, the diagnosis of AKI is easily overlooked by physicians in other divisions apart from nephrology. Yang reported that 74.2% of AKI patients could not be recognized in time during hospitalization [10].

Early identification of patients with high-risk factors enables a proactive strategy for AKI prevention and survival improvement [6]. Several predictive models of AKI occurrence has been developed by logistic regression based on independent variables in different clinical settings [7-10]. Yet, in reality, the risk factors of AKI are often codependent with complex interrelationships. It does not meet the independent assumptions of logistic regression, which could lead to wrong inferences. Hence, developing a more flexible and efficient AKI prediction algorithm will be a suitable strategy for early recognition of AKI. As a machine-learning algorithm, Bayesian networks (BNs) produce an intuitive, transparent, graphical representation of the interrelationships between factors, and can accurately reveal potential overall information [11]. The non-strict requirements for statistical assumptions in BNs modeling also made it of special significance in epidemiological studies [12]. Least absolute shrinkage and selection operator (LASSO) regression is regarded as an optimal variable selection method and is especially suitable for processing multi-collinear data or high-dimensional data. Previous studies proved that inserting LASSO regression into BNs analysis not only helps simplify the complexity of the network, but also improve the predictive accuracy of the model [13, 14].

To this end, this study aims to explore Bayesian networks to delineate the interrelationships between AKI and its associated risk factors in HM patients, to assess the prediction ability of BNs model, and then to infer the probability of AKI under different clinical settings.

Methods

Study design and participants

This study was designed as a retrospective cohort study based on the electronic medical records in Zhongshan Hospital, Fudan University, Shanghai of China. Hospitalized patients during Oct. 1st, 2014 and Sept. 30th, 2015 with the diagnosis of hematologic malignancies, including lymphoma, leukemia and multiple myeloma, were recruited. The exclusion criteria were: (1) with a stay of admission less than 24 hours; (2) with estimated glomerular filtration rate (eGFR) less than 30 mL/min per 1.73 m² [15, 16]; (3) on any forms of renal replacement therapy (RRT); (4) with only one test of SCr. If a patient was hospitalized multiple times during the study period, we regarded each hospitalization as an independent case.

Data collection

Data on age, gender, admission condition and diagnosis, pre-existing comorbidities and therapeutic regimen was exported and collected from the electronic medical records in Zhongshan Hospital. Biochemical indicators were exported from the lab database, and the baseline levels were set as the first test within 24 hours once admission. They were divided into three blocks: (1) Liver function: alanine aminotransferase (ALT), aspartate aminotransferase (AST) and total bilirubin (TBil); (2) Renal function: SCr, eGFR and serum uric acid (SUA); (3) Other biochemical indicators: albumin, hemoglobin, white blood cell (WBC) count and serum sodium and potassium. For patients in absence of SCr test on admission, the

outpatient SCr record in the past three months, if possible, was retrieved to estimate the baseline values. For those receiving over 2 times of SCr tests, we regarded the minimum SCr as baseline and the highest one as the peak value. The difference was calculated between the peak and baseline SCr values within seven days.

Definition and classification

In accordance with the 2012 Kidney Disease: Improving Global Outcomes guideline [17], AKI is defined as an absolute increase in SCr by ≥ 0.3 mg/dL within 48 hours or ≥ 1.5 -fold from the baseline within 7 days. Since the urine output cannot be accurately dated from the system, we only adopted the creatinine criteria for AKI diagnosis. The severity of AKI was categorized as Stage 1: increase in SCr ≥ 0.3 mg/dL or ≥ 1.5 -fold to 1.9-fold baseline; Stage 2: increase in SCr ≥ 2.9 –3.0 fold baseline; Stage 3: increase in SCr ≥ 3.0 fold baseline, ≥ 4.0 mg/dl increase or initiation of RRT. On the basis of International Classification of Diseases, 10th Revision (ICD-10), hematologic malignancy was categorized as lymphoma (C91-C95), leukemia (C81-85) and multiple myeloma (C90) in this study [18]. Anti-tumor treatment was divided into autologous stem cell transplantation (ASCT), chemotherapy and untreated/palliative care. Comorbidities were determined by the diagnosis records on admission and discharge. The reference value range of serum sodium, potassium, in this study was 137 to 147 mmol/L and 3.5 to 5.3 mmol/L. Values lower/higher than the reference range were considered as hypo-/hypernatremia and hypo-/hyperkalemia. The reference range for eGFR (based on EPI formula) and SUA were ≥ 90 mL/min/1.73m² and ≤ 359 μ mol/L, respectively. Anemia refers to hemoglobin <115 g/L and hypoalbuminemia refers to albumin <35 g/L.

Group LASSO regression

The LASSO procedure is a shrinkage method within least square method that enables to shrink estimation of irrelevant variables towards zero, allowing for automatic variable selection [19]. Group LASSO (gLASSO) is an extension of the LASSO to do the grouped variable selection in regression models based on predefined sets of variables. The estimates have the attractive property of being invariant under groupwise orthogonal reparameterizations. Assuming that we have J groups of categorical variables $\{G_1, G_2, \dots, G_J\}$ and each of them had $p_1, p_2, \dots, p_J$ levels, the gLASSO estimator is presented as: (see Formula 1 in the Supplementary files)

Through adjusting penalty l_1 and l_2 it can select candidate variables in group level and keep them invariant under group orthogonal transformation like ridge regression – i.e., within a group, coefficients will either all be zero or all nonzero. The penalty functions of *grLasso*, *grMCP*, and *grSCAD* carry out group selection, while the *gel* and *cMCP* penalties carry out bi-level selection. The point estimation of fitted lambda (λ) along the regularization path is selected according to *AIC*, *BIC*, or *GCV* criteria. Then, k-fold cross-validation for penalized gLASSO models should be performed to plot a grid of values for the regularization parameter λ . The lambda.min refers to the optimal variable selection with the minimum cross-validation error. Compared with the logistic model, gLASSO is more suitable for multi-collinear or

high-dimensional data. Even if the number of predictors is much larger than the sample size, the gLASSO estimator is shown to be statistically consistent.

Bayesian networks

A Bayesian network is a directed acyclic graph based on probability theory and graph theory, which consists of nodes representing the variables $U=\{X_1, \dots, X_n\}$ and directed edges symbolizing the relationships between the variables. If there is an edge from X_i to X_j , then we define that node X_i is the parent of X_j , and X_j is the child of X_i . Each node has a conditional probability distribution table, which quantitatively describes the probability dependence of the nodes and its parent nodes. From the perspective of probability theory, a BN represents the joint distribution of a set of random variables, according to the chain rule and conditional independence, the joint distribution of a series of random variables $X=\{X_1, \dots, X_n\}$ can be written as: (see Formula 2 in the Supplementary files)

$\pi(X_i)$ is the collection of the parent of X_i , $\pi(X_i) \in \{X_1, \dots, X_{i-1}\}$, given the value of $\pi(X_i)$ is conditionally independent of other variables in $\{X_1, \dots, X_{i-1}\}$. BNs modeling is usually performed as a two-step process: structure learning (learning the graph structure from the data) and parameter learning (learning the local distributions implied by the graph structure learned in the previous step). The structure learning can be traced to three approaches: constraint-based algorithms, score-based algorithms and hybrid algorithms. While the maximum likelihood estimation or bayesian estimation method is used to computing the probability of each node in the network conditions. BNs inference refers to calculating the posterior probability of X in the case of new evidence E . When E changes, conditional probability distributions of both parent and child nodes are affected. There are two algorithms for BNs inference, a logical sampling algorithm and a likelihood weighting algorithm, and the latter has a lower variance.

Statistical analysis

Categorical variables were summarized as proportions, and compared by using Pearson test and Cochran-Mantel-Haenszel (CMH) test. We estimated the crude odds ratios (cOR) and its 95% confidence interval (CI) of associated factors of AKI, including comorbidity, anti-tumor treatment, liver/renal function and biochemical test. Statistical description and univariate analysis were performed in IBM SPSS 22.0 (IBM Corp., Armonk, NY, USA). Significance for all statistical tests was a priori at $p<0.05$ and all p-values were two-tailed. Variable selection was conducted by gLASSO regression: firstly, these category variables were decomposed as dummy variables and their group label was assigned into another parallel dataset; the dummy and group datasets were analyzed in “grpreg” packages of R program 3.6.0 (R core team); *grLasso* penalty and *BIC* criteria was set to estimate the fitted lambda (λ) and 10-fold cross-validation was performed to screening the optimal variable selection with the minimum cross-validation error. Then, the selected predictors further created a Bayesian network in “bnlearn” packages in R program. The tabu-search algorithm was chosen to establish the BNs structure, and the maximum likelihood method was used to acquire the parameters for conditional probability distribution. The area under the receiver operating characteristic curve (AUC) was defined to assess the prediction ability of the BNs model. A 10-

fold cross-validation also was performed to reduce overfitting bias and for internal validation. Model drawing was done in Netica 5.18 (Norsys Software Corp., Vancouver, BC, Canada). Weka 3.8.0 (Waikato Environment for Knowledge Analysis, the University of Waikato, New Zealand) was used for evaluating the prediction ability of BNs models.

Results

Of the 2501 patients with hematologic malignancies, 2395 eligible participants were recruited in this study (Supplement Figure 1). The average age was 54.9 ± 15.5 years old and male patients accounted for 57.4% (n=1375).

AKI incidence and risk factors

A total of 370 (15.4%) HM patients were diagnosed with AKI during hospitalization. Of them, 308(12.9%), 41(1.7%) and 21(0.9%) episodes was in AKI stage 1, 2 and 3, respectively. The number of patients required renal replacement therapy was twenty (5.4%). Stratified by HM category, patients with multiple myeloma (24.1%) and leukemia (23.9%) shared a higher AKI incidence than lymphoma (13.4%).

Discreting age into four groups, patients under 29 years old had the highest risk of AKI (cOR: 2.16). The AKI incidence in female patients was higher than in the male (18.2% vs. 13.4%). Pre-existing diabetes increased the risk of AKI, while the difference was not found significant in hypertension. Compared with untreated/palliative care, ASCT and chemotherapy were the major contributors related to AKI with a cOR value of 4.37 and 2.24 respectively. Another significant association between risk factors and AKI were observed in liver and renal dysfunction: patients with abnormal ALT, AST and SCr values on admission were more predisposed to AKI; insufficient eGFR and increased SUA level also rose the probability of developing AKI. Patients who had initial decreased levels of albumin and hemoglobin were more venerable to AKI (cOR=2.72 and 3.85), so were patients with electrolyte abnormities (Table 1).

Variable selection in gLASSO

The tuning parameter (λ) was designated in gLASSO regression by using 10-fold cross-validation in Figure 1A. The vertical lines were drawn at the optimal λ value by using *grrasso* penalty. When setting $\log(\lambda)$ as -4.529, eleven variables were selected out of the initial nineteen variables with the minimum cross-validation error. It included age, gender, diabetes, HM category, anti-tumor treatment, hemoglobin, SCr, eGFR, SUA, serum sodium and potassium levels. Figure 1B presented the gLASSO coefficient (β) profiles of candidate variables. Similar 11 predictors with nonzero coefficients were identified when the *gLASSO* model met *BIC* criteria, where the fitted λ was estimated at 0.00896.

Bayesian network model of HM-related AKI

A probabilistic model with 19 the selected predictors and AKI was built by using BNs analysis Directed edges represent probabilistic dependencies between the nodes that are connected rather than the causal relationship. Figure 2 presented that correlations between AKI and its predictable factors were established

by a complex network structure, in which a direct connection between age, hemoglobin, eGFR, serum sodium and potassium and AKI was found. In addition, HM category and anti-tumor treatment were indirectly linked to AKI through hemoglobin and eGFR, while diabetes created a connection with AKI through affecting sodium levels. Interrelationship between the covariates was also presented that hemoglobin was associated with gender, HM category and anti-tumor treatment; eGFR was linked directly with age, HM category, anti-tumor treatment and SUA level. The marginal probability of AKI was estimated at 16.5%. The conditional probability distribution table of AKI was listed in Table 2, which quantitatively describes the relationship between AKI and its parent nodes of eGFR, hemoglobin and serum sodium. Patients whose eGFR was under 59 mL/min per 1.73 m² combined together with anemia and hyponatremia shared the highest AKI incidence (78.5%). In a similar situation but hypernatremia, the probability of AKI was estimated as 68.3%. In contrast, the lowest rate (5.2%) lay in patients with normal eGFR, hemoglobin and sodium level.

Bayesian network evaluation and model inference

As shown in Figure 3, the AUC value of BNs model was 0.835 (95% CI: 0.812 to 0.858) and higher than that of the logistic score model (0.763). Applying 10-fold cross-validation for evaluation, the AUC maintained at a level of 0.812 (95% CI: 0.787 to 0.837). By using Mantel-Haenszel test, no statistical significance of the difference in predictive accuracy was found between initial and cross-validation datasets ($p=0.298$). Based on the patient's demographic characteristics, medical history and clinical parameters, BNs could infer the probability of AKI individually during hospitalization. For instance, when a patient suffered from leukemia is initially detected with anemia, hyperuricemia and hyponatremia, the expected probability of subsequent AKI is estimated as 53.8% by a priori information about the BNs. However, once these biochemical indicators could be corrected to normal level timely, the risk of AKI is reduced to 9.9% (Figure 4).

Discussion

With the development of novel chemotherapeutic agents and targeted medicine, the survival time and quality of life have been remarkably improved among cancer patients. Meanwhile, the periodic anti-tumor treatment also poses them a higher risk of liver and renal dysfunction [20]. In this study, the incidences of AKI in patients with multiple myeloma, leukemia and lymphoma were estimated as 24.1%, 23.9% and 13.4% respectively. It is higher than in general inpatients [21-23] and patients with solid tumor [24, 25]. Hence, it is essential to develop novel models to predict the probability of AKI so as to take measures to better prevent AKI and adverse consequences related to deteriorate renal function during anti-tumor therapy.

In traditional logical regression, model predictions can only be performed when the state of all variables is known. From a health economic perspective, it is impractical to require patients to undergo all the tests included in the model. Thus, flexible disease prediction based on incomplete data may be more in line with clinical practice. In BNs model, probabilistic inference can be conducted based on the partially

available information rather than all of them. The parameters of unknown variables are estimated by using the prior knowledge acquired from BNs modeling. It enables physicians to infer the patients' individual AKI risk flexibly and easily. With the accumulation of patient information, the structure and parameters of BNs model will be further continuously corrected, and its predictive accuracy will also be improved. In the present study, we also found that the AUC value of AKI-BNs model was estimated at 0.835, which is higher than that of logistic score model (0.763). Robust performance in 10-fold cross-validation also indicates its suitability of clinical promotion.

It was observed that the occurrence HM-related AKI is usually multifactorial, not only due to pre-existing comorbidities, liver/renal dysfunction, anemia, but also the potential malignancies itself and anti-tumor treatment. Complex interrelationships between AKI and these risk factors makes it inadaptably to analyze in logistic methods. Multicollinearity among variables is commonly encountered in clinical data analysis and should be carefully considered unless it may lead to incorrect inferences. Penalization and regularization techniques have been proved to be the best algorithm to reduce the complexity of high-dimensional data. LASSO is one of the most recognized approaches to regularized regressions, which is suitable for analyzing a large number of clinical factors and avoiding overfitting [26]. As the extension of LASSO method, gLASSO can implement grouped variable selection and overcomes the limitations of LASSO to select only single variables. To this end, we applied gLASSO regression to screen out the 11 main predictors, and then constructed the BNs model to estimate conditional probability distribution for each node. Variable selection before modeling can help simplify the network structure, and guarantee the causal rationality. Nowadays, LASSO is widely applied as an effective tool for variable selection in BNs and other machine learning modeling [27, 28].

Our results revealed that age, hemoglobin, eGFR, serum sodium and potassium were directly linked with AKI. HM category and AKI was linked indirectly with hemoglobin and eGFR. When occurring in patients with chronic kidney disease (CKD), AKI will be more severe and difficult to recover. CKD is characterized by chronic inflammation and vascular dysfunction. These pathologic changes may contribute to the heightened susceptibility of AKI and partial even non-recovery. Anemia is also a common feature in HM patients, which can be caused by reduced hematopoietic capacity of bone marrow, hemodilution, frequent blood sampling, dysfunctional iron metabolism, nutritional deficiencies, diminished red blood cell survival, and a blunted cellular response to erythropoietin et al. One respective study in Korea also reports that anemia was more prevalent in HM patients than those with solid tumors (79.4% vs. 50.4%), and HM patients also share a higher risk of AKI and long-term mortality [29].

Apart from the conventional risk factors, our study reveals that electrolyte disturbance was also associated with a higher risk of AKI. *Olgar et al.* also found that in leukemia patients, 11.7% and 9.5% had hyponatremia and hypernatremia, 7.6% and 6.0% had hypokalemia and hyperkalemia [30]. Volume depletion is the main reason for hyponatremia, usually seen in hemorrhage, diarrhea, vomiting, drainage of ascites etc., which are common in HM patients receiving chemotherapy. Poor nutrition, anorexia and volume depletion can induce the inadequate potassium intake to cause hypokalemia. In patients with leukemia, the increased production of blast cells is also reported leading to hypokalemia [31]. Consistent

with our study, the HM category is recognized to cast an influence on renal insufficiency, mainly manifestes as reduced eGFR or increased SCr [2]. Lymphomatous or leukemic infiltration can lead to enlarged kidneys. Leukemic hyperleukocytosis with microcapillary obstruction and renal vein thrombosis alters renal vascular permeability. Lymphadenopathy as well as drug-induced crystalluria (acyclovir, cotrimoxazole) can result in obstructive nephropathy. Moreover, ASCT was observed with a higher risk of AKI, involving potential mechanisms including the adverse effect of calcineurin inhibitors, graft versus host disease and hepatic sinusoidal obstruction syndrome, which can induce anemia and renal dysfunction[32].

According to a national report from the United Kingdom, AKI can be avoided in one-fifth of patients who developed after admission, if they had better monitoring of electrolytes, recognition of risk factors, and prompt management [33]. In this study, the established BNs model inferred the probability of AKI, which made it possible to recognize patients with high risks on admission and guide the subsequent prophylaxis. When a patient suffered from leukemia was initially detected as anemia, hyperuricemia and hyponatremia, the expected probability of AKI was estimated as 53.8% by a priori information about the BNs. If these biochemical indicators were corrected to normal level timely, the incidence of AKI reduced to 9.9% notably. Hence, the BNs model has a flexible inference mechanism, which is beneficial to disease prediction and clinical diagnosis.

Our study is the first application of BNs in AKI study field. The visual network represents graphical interactions between risk factors and AKI, and can accurately reveal potential overall information. BNs also provides us with a new perspective that these risk factors may not always directly cause the occurrence of AKI, but rather indirectly through intermediate variables. The study was limited in certain aspects. Firstly, this was a single center-based study, which might restrict the extrapolation of research findings. Secondly, we did not collect the medication history of nephrotoxic drugs and it would obscure the relationships between chemotherapy and AKI incidence. Thirdly, our study was based on retrospective clinical data. BNs only revealed the relevant AKI risk factors, and the edges between variables only represent probability dependencies, not causal relationships. Therefore, the causal inference needs to be further verified in a prospective cohort together with the consideration of professional knowledge as well.

Conclusion

AKI is common in patients with HM, influenced by a variety of factors including comorbidity, renal/liver dysfunction and anti-tumor treatment. Bayesian networks combines with gLASSO can analyze not only how the correlative factors affect AKI but also their interrelationships. Inference using BNs model enables physicians to identify patients with a higher risk of AKI and implement prophylaxis and other protective strategies to improve clinical outcome.

Abbreviations

AKI: Acute kidney injury; ALT: Alanine aminotransferase; ASCT: Autologous stem cell transplantation; AST: Aspartate aminotransferase; AUC: The area under the receiver operating characteristic curve; BNs: Bayesian networks; CI: Confidence interval; CKD: Chronic kidney disease; cOR: crude Odds ratios; eGFR: estimated Glomerular filtration rate; gLASSO: group Least absolute shrinkage and selection operator; HM: Hematologic malignancies; LASSO: Least absolute shrinkage and selection operator; RRT: Renal replacement therapy; SCr: Serum creatinine; SUA: Serum uric acid; TBil: Total bilirubin; WBC: White blood cell;

Declarations

Ethics approval and consent to participate

This study was approved by the institutional review board of Zhongshan Hospital, Fudan University. Subjects was informed and recruited voluntarily. Patients' identification information was replaced with codes before data extraction for privacy concerns.

Consent for publication

The Author confirms that its publication has been approved by all co-authors.

Availability of data and materials

Data can be available by contacting the corresponding author.

Competing interests

The authors have no conflicts of interest to declare.

Funding

The study was financially supported by the Major Projects of Scientific Research, Innovation Plan of Shanghai Education Commission (no. 2017-01-07-00-07-E00009) and Shanghai Medical Center of Kidney (no. 2017ZZ01015) and Zhongshan Hospital Science Foundation for Youths (no.2019ZSQN19).

Author Contributions

YL participated in the study design, led the data analysis, and drafted the manuscript. XHC, YMW and JCH were involved in data collection and data analysis. ZYS edited the manuscript. XQD was responsible for this project and commented on the manuscript.

Acknowledgements

None.

References

1. Christiansen CF, Johansen MB, Langeberg WJ, Fryzek JP, Sørensen HT: Incidence of acute kidney injury in cancer patients: a Danish population-based cohort study. *European journal of internal medicine* 2011, 22(4):399-406.
2. Canet E, Vincent F, Darmon M, Soares M: Acute kidney injury in hematological patients. *Current opinion in critical care* 2015, 21(6):549-558.
3. Harris KP, Hattersley JM, Feehally J, Walls J: Acute renal failure associated with haematological malignancies: a review of 10 years experience. *European journal of haematology* 1991, 47(2):119-122.
4. Lahoti A, Nates JL, Wakefield CD, Price KJ, Salahudeen AK: Costs and Outcomes of Acute Kidney Injury in Critically Ill Patients with Cancer. *Journal of Supportive Oncology* 2011, 9(4):149-155.
5. Canet E, Zafrani L, Lambert J, Thieblemont C, Galicier L, Schnell D, Raffoux E, Lengline E, Chevret S, Darmon M *et al*: Acute kidney injury in patients with newly diagnosed high-grade hematological malignancies: impact on remission and survival. *PloS one* 2013, 8(2):e55870.
6. Wang Y, Fang Y, Teng J, Ding X: Acute Kidney Injury Epidemiology: From Recognition to Intervention. *Contributions to nephrology* 2016, 187:1-8.
7. Palomba H, de Castro I, Neto AL, Lage S, Yu L: Acute kidney injury prediction following elective cardiac surgery: AKICS Score. *Kidney Int* 2007, 72(5):624-631.
8. Jiang W, Teng J, Xu J, Shen B, Wang Y, Fang Y, Zou Z, Jin J, Zhuang Y, Liu L *et al*: Dynamic Predictive Scores for Cardiac Surgery-Associated Acute Kidney Injury. *Journal of the American Heart Association* 2016, 5(8).
9. Kim WH, Lee SM, Choi JW, Kim EH, Lee JH, Jung JW, Ahn JH, Sung KI, Kim CS, Cho HS: Simplified clinical risk score to predict acute kidney injury after aortic surgery. *Journal of Cardiothoracic & Vascular Anesthesia* 2013, 27(6):1158-1166.
10. Madhavan MV, Généreux P, Rubin J, Palmerini T, Caixeta A, Xu K, Weisz G, Mehran R, Stone GW: Usefulness of the SYNTAX Score to Predict Acute Kidney Injury After Percutaneous Coronary Intervention (from the Acute Catheterization and Urgent Intervention Triage Strategy Trial). *American Journal of Cardiology* 2014, 113(8):1331-1337.
11. Scutari M: Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* 2010, 35(03):2010.
12. Fuster-Parra P, Tauler P, Bennasar-Veny M, Ligeza A, Lopez-Gonzalez AA, Aguiló A: Bayesian network modeling: A case study of an epidemiologic system analysis of cardiovascular risk. *Computer methods and programs in biomedicine* 2016, 126:128-142.
13. McNally RJ, Mair P, Mugno BL, Riemann BC: Co-morbid obsessive-compulsive disorder and depression: a Bayesian network approach. *Psychological medicine* 2017, 47(7):1204-1214.
14. Zhang T, Ma Y, Xiao X, Lin Y, Zhang X, Yin F, Li X: Dynamic Bayesian network in infectious diseases surveillance: a simulation study. *Scientific reports* 2019, 9(1):10376.

15. Levey AS, de Jong PE, Coresh J, El NM, Astor BC, Matsushita K, Gansevoort RT, Kasiske BL, Eckardt KU: The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. *Kidney International* 2011, 80(1):17-28.
16. Xu X, Nie S, Liu Z, Chen C, Xu G, Zha Y, Qian J, Liu B, Han S, Xu A: Epidemiology and Clinical Correlates of AKI in Chinese Hospitalized Adults. *Clinical Journal of the American Society of Nephrology Cjasn* 2015, 10(9):1510.
17. Disease K, Outcomes IG: Acute Kidney Injury Work Group: KDIGO clinical practice guideline for acute kidney injury. *Kidney International Supplement* 2012, 2(1):1-138.
18. International Statistical Classification of Diseases and Related Health Problems 10th Revision, <https://icd.who.int/browse10/2016/en>, Accessed 5 Feb 2017.
19. Tibshirani R: Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society* 1996, 58(1):267-288.
20. Susantitaphong P, Cruz DN, Cerda J, Abulfaraj M, Alqahtani F, Koulouridis I, Jaber BL: World Incidence of AKI: A Meta-Analysis. *Clinical Journal of the American Society of Nephrology Cjasn* 2013, 8(9):1482-1493.
21. Fang Y, Ding X, Zhong Y, Zou J, Teng J, Tang Y, Lin J, Lin P: Acute kidney injury in a Chinese hospitalized population. *Blood Purification* 2010, 30(2):120-126.
22. Cheng X, Wu B, Liu Y, Mao H, Xing C: Incidence and diagnosis of Acute kidney injury in hospitalized adult patients: a retrospective observational study in a tertiary teaching Hospital in Southeast China. *BMC Nephrol* 2017, 18(1):203.
23. Pavkov ME, Harding JL, Burrows NR: Trends in Hospitalizations for Acute Kidney Injury - United States, 2000-2014. *MMWR Morbidity and mortality weekly report* 2018, 67(10):289-293.
24. Riffaut N, Moranne O, Hertig A, Hannedouche T, Couchoud C: Outcomes of acute kidney injury depend on initial clinical features: a national French cohort study. *Nephrol Dial Transplant* 2018.
25. Salahudeen AK, Doshi SM, Pawar T, Nowshad G, Lahoti A, Shah P: Incidence Rate, Clinical Correlates, and Outcomes of AKI in Patients Admitted to a Comprehensive Cancer Center. *Clinical Journal of the American Society of Nephrology Cjasn* 2013, 8(3):347-354.
26. Hepp T, Schmid M, Gefeller O, Waldmann E, Mayr A: Approaches to Regularized Regression - A Comparison between Gradient Boosting and the Lasso. *Methods of information in medicine* 2016, 55(5):422-430.
27. Goto T, Camargo CA, Jr., Faridi MK, Yun BJ, Hasegawa K: Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *The American journal of emergency medicine* 2018, 36(9):1650-1654.
28. Xu H, Zhao X, Shi Y, Li X, Qian Y, Zou J, Yi H, Huang H, Guan J, Yin S: Development and validation of a simple-to-use clinical nomogram for predicting obstructive sleep apnea. *BMC pulmonary medicine* 2019, 19(1):18.
29. Han SS, Baek SH, Ahn SY, Chin HJ, Na KY, Chae DW, Kim S: Anemia Is a Risk Factor for Acute Kidney Injury and Long-Term Mortality in Critically Ill Patients. *The Tohoku journal of experimental medicine*

2015, 237(4):287-295.

30. Olgar S, Yetgin S, Cetin M, Aras T, Akhan O: Electrolyte abnormalities at diagnosis of acute lymphocytic leukemia may be a clue for renal damage in long-term period. *J Pediatr Hematol Oncol* 2005, 27(4):202-206.
31. Adams PC, Woodhouse KW, Adela M, Parnham A: Exaggerated hypokalaemia in acute myeloid leukaemia. *British medical journal (Clinical research ed)* 1981, 282(6269):1034-1035.
32. Andronesi AG, Tanase AD, Sorohan BM, Craciun OG, Stefan L, Varady Z, Lipan L, Obrisca B, Truica A, Ismail G: Incidence and risk factors for acute kidney injury following autologous stem cell transplantation for multiple myeloma. *Cancer medicine* 2019, 8(6):3278-3285.
33. Mayor S: UK report into acute kidney injury deaths urges electrolyte checks in all emergency admissions. *Bmj British Medical Journal* 2009, 338(7708):1407-1407.

Tables

Table 1. Associated factors of AKI in patients with hematologic malignancies

Variate	Total	AKI (%)	Z	p-value	cOR (95%CI)
Age					
<29 yr	213	57(26.8)	0.564	0.453*	2.16(1.47~3.18)
30~49 yr	539	78(14.5)			1.00
50~69 yr	1236	172(13.9)			0.96(0.72~1.28)
≥70 yr	407	63(15.5)			1.08(0.76~1.55)
Gender					
Male	1375	184(13.4)	10.561	0.001 [#]	1.00
Female	1020	186(18.2)			1.44(1.16~1.80)
Comorbidities					
Hypertension	473	75(15.9)	0.075	0.784 [#]	1.04(0.79~1.37)
Diabetes	814	159(19.5)	15.748	<0.001 [#]	1.58(1.26~1.98)
HM category					
Lymphoma	1941	261(13.4)	31.429	<0.001 [#]	1.00
Leukemia	201	48(23.9)			2.02(1.42~2.86)
Multiple Myeloma	253	61(24.1)			2.05(1.49~2.81)
In-hospital Condition					
Emergent	163	41(25.2)	12.610	<0.001 [#]	1.94(1.34~2.82)
Normal	2232	329(14.7)			1.00
Anti-tumor Treatment					
ASCT	50	13(26.0)	20.846	<0.001 [#]	4.37(2.04~9.36)
Chemotherapy	2036	334(16.4)			2.24(1.57~3.79)
Untreated/palliative	309	23(7.4)			1.00
Liver Function					
ALT (≥40 U/L)	150	35(23.3)	7.616	<0.001 [#]	1.74(1.17~2.58)
AST (≥35 U/L)	270	62(23.0)	13.154	<0.001 [#]	1.76(1.29~2.39)
TBiL(≥20.4 μmol/L)	90	17(18.9)	0.847	0.357 [#]	1.29(0.75~2.21)
Renal Function					
SCr(≥115 μmol/L)	113	79(69.9)	269.309	<0.001 [#]	15.9(10.44~24.21)
eGFR(≥90 mL/min/1.73m ²)	1650	170(10.3)	207.618	<0.001*	1.00
eGFR(60~89 mL/min/1.73m ²)	596	110(18.4)			1.96(1.51~2.55)
eGFR(≤59 mL/min/1.73m ²)	147	90(61.2)			13.75(9.52~19.86)
SUA(≤359 μmol/L)	1619	198(12.2)	88.388	<0.001*	1.00
SUA (360~420 μmol/L)	398	58(14.6)			1.22(0.89~1.68)
SUA (421~480 μmol/L)	200	40(20.0)			1.79(1.23~2.62)
SUA (≥481 μmol/L)	178	74(41.6)			5.11(3.66~7.13)
Biochemical Test					
Album (<35 g/L)	584	156(26.7)	75.012	<0.001 [#]	2.72(2.16~3.43)
Hemoglobin (<115 g/L)	1338	297(22.2)	105.701	<0.001 [#]	3.85(2.93~5.04)
WBC (≥9.5×10 ⁹)	406	102(25.1)	35.028	<0.001 [#]	2.15(1.66~2.79)
Hyponatremia	325	135(41.5)	216.335	<0.001 [#]	5.84(4.50~7.59)
Hypernatremia	41	15(36.6)			4.74(2.47~9.09)
Hypokalemia	333	122(36.6)	227.450	<0.001 [#]	4.59(3.53~5.97)
Hyperkalemia	24	20(83.3)			39.69(13.45~117.15)

* Cochran-Mantel-Haenszel(CMH) test; # Pearson Test;

AKI: acute kidney injury; cOR: crude odds ratio; HM: hematologic malignancy; ASCT: autologous stem cell transplantation; ALT: alanine aminotransferase; AST: aspartate aminotransferase; TBil: total bilirubin; SCr: serum creatinine; eGFR: estimated glomerular filtration rate; SUA: serum uric acid; WBC: white blood cell.

Due to technical limitations Table 2 is available in the Supplementary Files.

Supplemental Figure Legend

Supplement Fig. 1. Flow chart of the study population selection

Figures

Figure 1

AKI variable selection by using gLASSO regression

HM: hematologic malignancy; SCr: Serum creatinine; eGFR: estimated Glomerular filtration rate;

Figure 2

Bayesian Network model of factors relating to AKI in patients with HM

Figure 3

Receiver operating characteristic curves for AKI predictors in Bayesian network

HM: hematologic malignancy; SCr: Serum creatinine; eGFR: estimated Glomerular filtration rate;

Figure 4

Bayesian network under known evidence variables

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Formula1.jpg](#)
- [Table2.pdf](#)
- [Formula2.jpg](#)
- [FigureS1revised1127.docx](#)