Sex Differences in AKI Risk and Outcomes - Insights from MIMIC-III and MIMIC-IV

Rui Feng[1*] and Linge Lass[2]

[1*]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, 423 Guardian Dr, Philadelphia, 19104, PA, USA.
[2]Quantitative Research + Machine Learning, Susquehanna International Group, 401 E City Ave, Bala Cynwyd, 19004, PA, USA.

*Corresponding author(s). E-mail(s): ruifeng@upenn.edu;
Contributing authors: linge.lass@gmail.com;

Abstract

Purpose: Acute Kidney Injury (AKI) affects 13-18% of hospitalized patients in the United States, with men showing higher incidence than women and worse outcomes, including greater progression to Chronic Kidney Disease (CKD) and higher post-AKI mortality. The mechanisms underlying these sex differences remain poorly understood. This study aimed to develop machine learning models for predicting AKI, CKD after AKI, and post-AKI mortality, and identify common and sex-specific risk factors.

Methods: We conducted a secondary analysis of the MIMIC-III (2001-2012) and MIMIC-IV (2008-2019) critical care databases, including adults $\geq$ 18 years with complete hospitalization, demographic, and laboratory data. Predictors included demographics, comorbidities, vital signs, laboratory values, medications, interventions, and free-text clinical notes. We developed prediction models in MIMIC-IV and validated in MIMIC-III. Model performance from using ElasticNet logistic, Random Forest, XGBoost, DNN, DCN, and Transformer, using AUC-ROC, accuracy, precision, recall, and F1 score. We used regression coefficients and H statistic to quantify sex-predictor interaction effect on AKI. Competing risk models and deepSurv were used for modeling time from AKI to CKD and death.

Results: XGBoost outperformed other models in AUC-ROC and prediction metrics for AKI. In logistic models, high serum creatinine showed a stronger association with AKI in men (OR = 4.14 per 100-unit increase) than in women (OR = 1.37). Liver disease was a strong risk in both sexes with modest attenuation in men (ROR = 0.82). Other sex modifications include bicarbonate, hemoglobin, and widowed status. H-statistics from the random forest suggested additional

sex–predictor interactions, including age, private insurance, single marital status, Black race, vascular comorbidities, and creatinine. For CKD after AKI, hazards increased with age, black race, unmarried status, cardiovascular comorbidities, and higher potassium in both sexes. CKD was associated with CKD in men with lower pH, higher sodium, and higher creatinine, but not in women. For the competing event death, WBC, potassium and creatine were identified biomarkers, heart failure showed a greater hazard in woman than men.
Conclusion: We identified common and sex-dependent risk factors for AKI and CKD. Deep learning models improved prediction of AKI and CKD and may help identify more nonlinear, sex-dependent risk factors than classic regression models. The results support sex-specific risk stratification and targeted prevention to reduce sex differences in AKI and consequent CKD risk.

Keywords: Acute Kidney Injury, Sex Differences, Interaction Effect, Deep Learning

1 Introduction

Acute Kidney Injury (AKI) is a critical and often overlooked condition, affecting approximately 13% to 18% of hospitalized patients and a total of 1.5 to 2 million individuals annually in the United States [1, 2]. Men are at significantly higher risk, with an incidence rate of 15% to 20% in males compared to 10% to 15% in females in hospital settings [3].

Furthermore, men are more likely to experience severe types of AKI, associated complications, and death. Approximately 50% of men who suffer AKI may develop Chronic Kidney Disease (CKD) within five years, while the incidence in women is somewhat lower, around 40% [2, 4]. The mortality rate following AKI was approximately 30% in men compared to 20% for women. The reasons for these differences may include a higher prevalence of comorbidity like hypertension and diabetes in men, which exacerbate kidney damage [4, 5].

Despite these disparities, the mechanisms underlying differences in incidence and outcomes remain poorly characterized, calling for further investigation [6]. The Medical Information Mart for Intensive Care (MIMIC) database provides a rich resource of clinical data, enabling researchers to develop and validate predictive models for AKI, including existing models that provide a foundational understanding of overall AKI risk [7]. Leveraging this database, this study investigated sex differences in AKI, developed predictive models and the identification of common and unique risk factors associated with AKI and Chronic Kidney Disease (CKD).

Recent years have witnessed a surge in the application of machine learning (ML) models to predict acute kidney injury (AKI). A systematic review of 46 studies found that most efforts focused on intensive care populations, with the MIMIC database frequently used as the primary data source [8]. The models varied widely, ranging from flexible methods such as random forests and gradient-boosted trees to deep learning models like recurrent neural networks. Reported predictive performance spanned a broad range, with AUROC values between 0.49 and 0.99, highlighting both the potential and variability of these approaches. Many of these studies had methodological

limitations, including high risk for bias, lack of calibration, inadequate handling of missing data, and scarce external validation. Importantly, interpretability and clinical applicability were rarely emphasized, limiting translational value. Despite these gaps, the review underscored the potential of advanced ML methods—particularly when incorporating temporal data and unstructured text—to improve AKI prediction beyond traditional regression approaches.

Although existing ML models have advanced risk prediction, they have largely overlooked sex as a biological and clinical modifier, leaving a critical gap in understanding and personalized risk stratification. Our work explicitly modeled sex-specific effects and interactions, with the goal of developing interpretable, clinically applicable tools for AKI, CKD, and mortality risk prediction.

2 Method

2.1 Data Description

We used MIMIC-III (2001–2012) and MIMIC-IV (2008–2019), de-identified ICU electronic health records from Beth Israel Deaconess Medical Center. MIMIC-III contains >50,000 adult ICU admissions across >38,000 patients; MIMIC-IV expands to >546,000 admissions for >364,000 patients with an updated schema and richer documentation. We also used the supplemental MIMIC-IV-ED module, which captures ¿400,000 emergency department encounters at the same institution from 2011–2019, aligned to MIMIC-IV data. All datasets were obtained from PhysioNet after completion of human-subjects training and a data use agreement. These databases provide extensive structured (demographics, vital signs, laboratory results, medications, procedures) and unstructured (clinical notes) data for prediction modeling of AKI and its outcomes [9, 10].

2.2 Primary Outcomes and Subject Selection

We included adult patients aged 18 years or older. Incidental AKI was identified by KDIGO serum creatinine criteria ((i) increase in sCr $\geq$ 0.3 mg/dL within 48 h, or (ii) increase to $\geq$ 1.5$\times$ baseline within 7 days) and ICD-9/10 codes (ICD-9 584.*, ICD-10 N17.*) during hospital stay, excluding patients with previous kidney transplant (ICD-9 V42.0; ICD-10 Z94.0), ESRD/dialysis dependence (ICD-9 585.6, V45.11, V56.*; ICD-10 N18.6, Z99.2), and pre-existing CKD (ICD-9 585.1–585.5, 585.9; ICD-10 N18.1–N18.5, N18.9). For patients with multiple ICU admissions, we retained the first admission with AKI for cases and the last admission without AKI for controls.

CKD, a major long-term adverse outcome following AKI, was identified using ICD-9/10 codes (ICD-9 585.1–585.5, 585.9; ICD-10 N18.1–N18.5, N18.9). Among patients with incidental AKI, CKD patients were identified with $\geq$ 90 days CKD diagnosis; For those without CKD, we flagged ESRD/dialysis/transplant admissions $\geq$ 90 days as post CKD. Time to CKD or post CKD was calculated based on admission times. Death and time to death were identified from discharge data, ED records, and date of deaths. For patients without CKD or post CKD or death, we required evidence

of follow-up and censored at the last observed admission. Thus individuals with no post-index encounters or with <90 days of follow-up were excluded.

2.3 Candidate Predictor Variables

We considered multi-modal predictors, including demographics, vital signs, laboratory results, medications, procedures, and unstructured clinical notes for AKI model (Table 1). To avoid post-outcome information leakage, we did not include variables that could only be observed after AKI onset, such as, AKI stage, in-hospital treatments, and length of stay. We also omitted composite scores (Charlson index, GCS, SAPS II, SOFA, etc.) because these scores are derived from raw measurements already included in analysis, are highly collinear with them, and would double count information. Instead, we assume that severity and comorbidity were represented directly by the underlying clinical variables.

To balance recency with availability, we applied clinically meaningful look-back windows before AKI: (i) laboratory tests were restricted to results obtained within the **7 days** before the AKI admission for cases or the most recent admission for non-AKI, which captures short- term physiology relevant to imminent risk while staying strictly pre-event; (ii) physiological (vital) signs were taken from the **48 hours** before admission to reflect acute status at presentation without incorporating in- hospital responses; and (iii) medications and procedures were extracted from the **6 months** prior to admission to summarize baseline exposures and recent interventions, avoiding stale history. All text features, when used, were limited to notes authored before the admission timestamp.

For CKD prediction, we all defined three pre-event windows: 0–90, 91–180, and 181–365 days before CKD onset (or censoring) to separate the short-, intermediate-, and long-term effects of laboratory and medicine effects. We excluded text inputs because of extremely unbalanced data.

Compared to literature [8, 11], our study incorporated a broader spectrum of data modalities, including nominal variables, longitudinal trajectories, and unstructured text fields, rather than relying primarily on static or cross-sectional or even post-event predictors. In addition, all predictors were restricted to the time window preceding the AKI or CKD diagnosis, to ensure temporal alignment between exposures and outcome. Such design reflects real-world clinical decision-making and reduces the risk of incorporating post-diagnosis information. These diverse data sources provided a more holistic representation of patient risk and enabled a more accurate and interpretable modeling for predicting AKI and its downstream outcomes.

2.4 Data Pre-processing

For unstructured clinical notes, we normalized the text (lowercasing, stopword removal, lemmatization), embedded sentences through SentenceTransformer (SBERT) [12, 13], which translate semantically similar texts to nearby vectors. The resulting concepts were represented as biomedical embeddings for downstream models.

Table 1 Candidate Predictors for AKI Prediction Models

Category (Modifiable)	**Variables**	**Time Window**	**MIMIC III/IV tables**
Demographics (No)	sex, age, race, ethnicity, marital status	adm'n	`patients`, `admissions`
Insurance (No)	Private; Medicare; Assistant (Medicaid, government, self-pay)	adm'n	`admissions`
Comorbidities (Indirect)	Hypertension, diabetes, cancer, heart failure, coronary artery disease, infection history, COPD, liver disease, cerebrovascular disease, obesity, CKD	adm'n	`diagnoses_icd`
Laboratory Tests (Partial)	white blood counts (WBC), hemoglobin, platelets, creatinine, blood urea nitrogen (BUN), albumin, sodium, potassium, bicarbonate, lactate, anion gap, base excess, INR, PTT, prothrombin time, glucose, ABGs (pH, PaO_2, $PaCO_2$)	7 days pre-ICU	`labevents` `chartevents`
Physiologic signs (Partial)	heart rate, SBP/DBP, MAP, respiratory rate, FiO_2, SpO_2, temperature, cardiac output, ccentral venous pressure (CVP), pulmonary artery pressures, tidal volume, minute ventilation; fluid input/output	48 hours pre-ICU	`chartevents`, `inputevents_mv`, `inputevents_cv`, `outputevents`, `icustays`
Medications (Partial)	Diuretics, non-steroidal anti-inflammatory drugs (NSAIDs), vasopressors, aminoglycosides, mannitol, colloid bolus	30-day & 6-month pre-ICU	`prescriptions`
Procedures (Partial)	mechanical ventilation, vasopressors, central line	30-day & 6-month pre-ICU	`inputevents_mv`, `inputevents_cv`, `procedureevents_mv`, `chartevents`
Text notes (No)	Nursing/physician notes, progress/discharge summaries, radiology, procedure, medication, admission notes	During hospitalization	`noteevents`

"Assistant" aggregates Medicaid, government, and self-pay. Time windows are relative to the index AKI admission or ICU admission as indicated. "Modifiable" denotes potential clinical intervenability of the exposure (not causality).

For categorical predictors (e.g., comorbidity status, surgery type), very rare categories (prevalence $< 1\%$) were collapsed into clinically similar groups to reduce sparsity and improve stability.

For repeated laboratory and physiological measurements, we condensed values within the pre-defined time windows into summary statistics,including, minimum, maximum, mean, standard deviation, most recent value, and measurement count, to retain most signals while reducing noise and irregular sampling effects.

For medications and procedures, we encoded exposure using two time windows: a short term of 0-30 days and a long term of 31-180 days before the admission. This was designed to differentiate acute exposures from subacute/chronic background use, to facilitate clinical meaningful interpretations.

For CKD prediction, we also computed window-specific laboratory means within each pre-defined time window (0-90, 90-180, 180-365 days), and created binary indicators for any medication or procedure exposure within each window. Candidate predictors included all demographics, insurance, baseline comorbidities, laboratory values, medication, and procedures, with sex-predictor interaction terms.

2.5 AKI Risk Prediction Model

We developed prediction models for AKI using multiple machine learning models, including *Random Forests* [14], *XGBoost* [15], a *deep neural network (DNN)* [16], the *Deep & Cross Network (DCN)* [17, 18], and a *Transformer* [19]. *Random Forests* aggregate many decorrelated trees without the need for extensive tree triming, capture nonlinear effects and interactions, and has shown to be robust to outliers and noisy features. *XGBoost* provides sparsity-aware gradient boosting with explicit regularization and missing-value handling via learned default directions. It scales well on large tabular datasets and captures nonlinearities and higher-order interactions that linear models miss. A *DNN* stacks fully connected layers and learns hierarchical feature representations that model complex, nonlinear relationships common in heterogeneous EHR features [16]. *DCN* combines an explicit cross network (learning bounded-degree feature crosses) with a deep MLP, efficiently capturing salient tabular interactions without enumerating all crosses [17, 18]. *Transformer*-based modeling uses self-attention to capture long-range sequence dependencies and cross-variable relations; with appropriate time encodings/masking, it accommodates irregular sampling and multi-modal inputs typical of MIMIC [19].

Compared to traditional regression models, these machine learning methods better accommodated collinear predictors, rare outcomes, and nonlinear relationships, and identified interactions that might otherwise be overlooked.

To benchmark against machine learning models, we fit logistic regression with Elastic Net regularization [20]. Because summary statistics of laboratory and physiological measures were highly collinear, we included only the pre-event means to reduce redundancy and stabilize coefficient estimates. Sex$\times$predictor multiplicative terms were included to estimate and interpret sex-specific associations directly.

Models were trained on MIMIC-IV with a 60%/20%/20% split for training/-validation/testing and externally validated on the independent MIMIC-III cohort. Performance was summarized by optimal accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) using the testing sets.

2.6 Post-AKI Outcome Prediction Model

In a competing-risks framework, we applied a Fine–Gray subdistribution hazards model to predict CKD risk after AKI, treating death as the competing event. Because post-CKD events were rare, we considered them as late surrogates and merged them with CKD. The candidate predictors include all demographics, medical history, pre-AKI 7-day average biomarkers, pre-AKI mediciation and procedure, and their sex interactions. A step-wise model selection procedure using p¡0.05 was used to determine the final model.

We also explored a deep Cox model (DeepSurv) for time-to-event outcomes [21], including the baseline demographic and comorbidities, all short-, intermediate-, and long-term biomarker exposures before AKI.

We used MIMIC-IV as the training data and MIMIC-III as the testing set. Model performance for survival predictions was evaluated using time-dependent AUC-ROC ([22] and the Brier score ([23]).

2.7 Predictor Explainability and Sex-predictor Interaction

We quantified feature importance and contribution to the prediction models using three complementary families of methods: a model-specific *feature importance index* (FI), Shapley values (SHAP), and Integrated Gradients (IG). For tree ensembles (Random Forests and XGBoost), we computed gain-based importance from split improvements [14, 24–26]. For neural models (DNN, DCN, Transformer), we computed SHAP values and Integrated Gradient attributions with a path integral of $m = 50$ steps from a cohort-specific baseline; baselines were set to sex- and age-stratified (± 2 years) medians to preserve clinical plausibility [27–29]. We aggregated the importance measured over models and ranked them for comparisons.

We investigated sex–predictor interactions with methods that provide interpretability in addition to model performance.

In the regression models, we assessed interaction terms directly by estimating and testing whether predictor slopes differed by sex.

For the deep learning models, we used the H-statistic of Friedman and Popescu [30] to quantify interactions. This measure estimates the variance in the model's predictions that is attributable to the interaction between two predictors, relative to their individual contributions. Specifically, if two predictors do not interact, the joint partial dependence function can be decomposed into the sum of their individual functions centered at zero. The H-statistic captures the deviation from this additive structure, with values ranging from 0 (no interaction) to 1 (interaction explains all variation). In some cases, values may exceed 1 when interaction effects dominate over marginal effects.

By applying the H-statistic to sex–predictor pairs, we were able to isolate sex-specific contributions to risk prediction and identify variables with effects differing between men and women. Because H requires repeated PD evaluations and can be computationally heavy on large EHR data, we implemented a stratified sampling strategies to ensure scalability.

2.8 Sensitivity Analyses

We conducted sensitivity analyses to address missingness, feature availability, and case mix.

In the AKI models, variables with $>70\%$ missingness were excluded. For variables with $\leq 70\%$ missingness, we imputed values using sex- and age-stratified (within ± 2 years) medians as the primary approach. As a sensitivity check, we performed multiple imputation (10 repeats) using a sequential multiple imputation method ([31, 32]) assuming missing at random.

Because automobile accidents caused a large percentage of AKI and the subjects might have different clinical profiles from other AKI patients, we repeated AKI analyses after excluding automobile/transport accident encounters (e.g., ICD-9 E810–E829; ICD-10 V00–V89).

For CKD and mortality outcomes, we also considered 60-day windows for CKD incidences and deaths because of possibly delayed time stamp for AKIs.

3 Results

3.1 Study Cohort Characteristics

The characteristics of the study cohorts are shown in Table 1. In MIMIC-III, AKI accounted for 20.0% of eligible admissions (6,828/34,181); in MIMIC-IV, it was 13.1% (26,184/199,822). AKI patients were older (66.7±16.3 vs 62.1±17.8 years in MIMIC-III; 64.1±16.7 vs 52.2±20.0 years in MIMIC-IV), male-dominant (56.6% in MIMIC-III; 56.4% in MIMIC-IV), and more often Black (9.9% vs 6.1% in MIMIC-III; 13.2% vs 12.4% in MIMIC-IV). They were more frequently widowed (16.8% vs 12.6% in MIMIC-III; 13.1% vs 7.4% in MIMIC-IV) and divorced/separated (8.1% vs 6.7% in MIMIC-III; 7.4% vs 6.1% in MIMIC-IV). Insurance shifted toward Medicare and away from private coverage—Medicare: 61.4% vs 48.3% (MIMIC-III) and 54.5% vs 32.2% (MIMIC-IV); private: 25.9% vs 38.3% (MIMIC-III) and 25.8% vs 42.8% (MIMIC-IV).

Figure 1 (a) and (b) show the age-specific incidence rates of AKI stratified by sex, in MIMIC-III and MIMIC-IV data, respectively. As expected, males had higher AKI incidence than females across all age groups. The incidence rates in MIMIC-III, which contains ICU-only data, were substantially higher than in the general U.S. population. However, the relative risk (male vs. female) remained consistent across both MIMIC cohorts.

Table 2 Characteristics of Study Cohorts

Variable	MIMIC-III		MIMIC-IV	
	No AKI (N=27,353)	AKI (N=6,828)	No AKI (N=173,638)	AKI (N=26,184)
Age Mean±SD	62.1±17.8	66.7±16.3	52.2±20.0	64.1±16.7
Gender				
Female	12,049 (44.1%)	2,965 (43.4%)	96,865 (55.8%)	11,408 (43.6%)
Male	15,304 (56.0%)	3,863 (56.6%)	76,773 (44.2%)	14,776 (56.4%)
Race and Ethnicity				
White	19,471 (71.2%)	4,889 (71.6%)	114,777 (66.1%)	17,451 (66.6%)
Black	1,682 (6.1%)	679 (9.9%)	21,613 (12.4%)	3,455 (13.2%)
Hispanic	930 (3.4%)	185 (2.7%)	10,102 (5.8%)	1,067 (4.1%)
Asian	635 (2.3%)	171 (2.5%)	8,195 (4.7%)	816 (3.1%)
Others	3,652 (13.4%)	716 (10.5%)	10,258 (5.9%)	2,391 (9.1%)
Unknown	983 (3.6%)	188 (2.8%)	8,693 (5.0%)	1,004 (3.8%)
Marital Status				
Married	13,357 (48.8%)	3,072 (45.0%)	74,208 (42.7%)	11,019 (42.1%)
Single	6,707 (24.5%)	1,684 (24.7%)	68,665 (39.5%)	8,010 (30.6%)
Widowed	3,444 (12.6%)	1,144 (16.8%)	12,836 (7.4%)	3,420 (13.1%)
Divorced/Separated	1,847 (6.7%)	548 (8.1%)	10,648 (6.1%)	1,938 (7.4%)
Others/Unknown	1,998 (7.3%)	380 (5.6%)	7,281 (4.2%)	1,797 (6.9%)
Insurance				
Private	10,479 (38.3%)	1,767 (25.9%)	74,288 (42.8%)	6,763 (25.8%)
Medicare	13,222 (48.3%)	4,193 (61.4%)	55,831 (32.2%)	14,279 (54.5%)
Assistant	3,652 (13.4%)	868 (12.7%)	32,005 (18.4%)	4,174 (15.9%)
Unknown	0 (0.0%)	0 (0.0%)	11,514 (6.6%)	968 (3.7%)
Comorbidity				
Hypertension	13,652 (49.9%)	3,537 (51.8%)	63,895 (36.8%)	16,038 (61.3%)
Diabetes	5,562 (20.3%)	1,973 (28.9%)	22,849 (13.2%)	7,446 (28.4%)
Coronary artery disease	7,903 (28.9%)	1,918 (28.1%)	19,818 (11.4%)	6,818 (26.0%)
Heart failure	4,477 (16.4%)	2,372 (34.7%)	9,487 (5.5%)	5,836 (22.3%)
COPD	4,492 (16.4%)	1,340 (19.6%)	14,700 (8.5%)	4,143 (15.8%)
Cancer	3,219 (11.8%)	1,223 (17.9%)	14,509 (8.4%)	5,182 (19.8%)
Infection	1,945 (7.1%)	2,340 (34.3%)	15,600 (9.0%)	9,224 (35.2%)
Liver disease	1,307 (4.8%)	1,324 (19.4%)	6,646 (3.8%)	4,822 (18.4%)
Cerebrovascular disease	1,994 (7.3%)	482 (7.1%)	5,469 (3.1%)	1,539 (5.9%)
Obesity	1,213 (4.4%)	405 (5.9%)	13,085 (7.5%)	3,094 (11.8%)
AKI Outcomes				
All-cause death	-	4,104 (60.1%)	-	4,265 (16.3%)
CKD after 90 days	-	113 (1.7%)	-	2,990 (11.4%)
ESRD/TX/Dialysis	-	23 (0.3%)	-	188 (0.7%)

3.2 AKI Prediction Model

The ROC curves by different models are shown in Figure 2. XGboost achieved the highest AUC-ROC on both MIMIC-IV testing set (0.927) and MIMIC-III (0.829), which outperformed Logistic by +0.076 and +0.055, respectively. On MIMIC-IV testing set, the next best AUCs were Transformer (0.916), Random Forest (0.909), DNN (0.908), and DCN (0.904); on MIMIC-III they were Random Forest (0.777), Logistic (0.774), DNN (0.764), Transformer (0.739), and DCN (0.723). All models showed expected degradation under external validation, with AUC drops ranging 0.077–0.181; Logistic

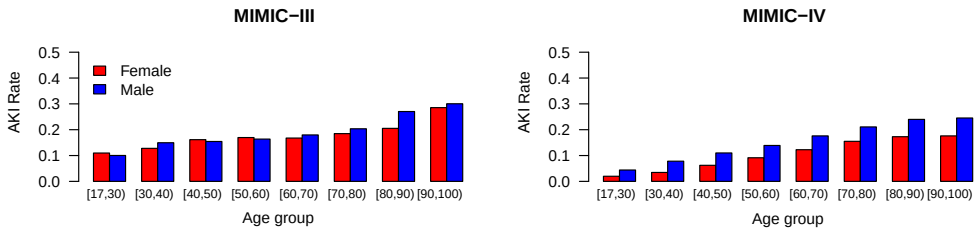

Fig. 1 AKI rates in age groups of males and females in MIMIC-III and MIMIC-IV

had the smallest decline (0.077), followed by XGboost (0.098), suggesting their generalization of simple models. As a reference logistic model with all demographic variables included, the ROCs were 0.696 and 0.581 in MIMIC-IV and MIMIC-III, respectively.

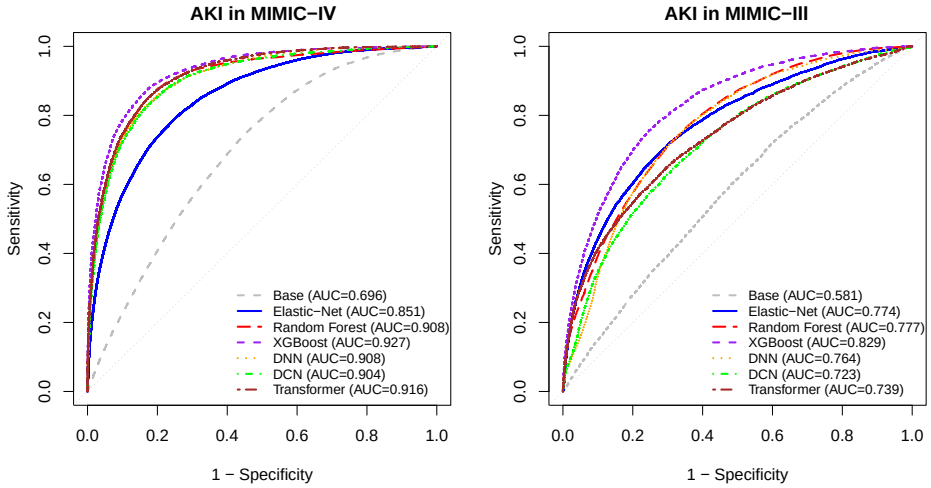

Fig. 2 ROC of Prediction models for AKI in MIMIC-III and MIMIC-IV

Other metrics (accuracy, precision/F1, sensitivity/specificity) broadly tracked the AUC rankings: XGboost maintained the best overall balance, with modest precision across models due to case-control imbalance (Table A).

3.3 Sex-Interaction Effects for AKI

Table 3.3 shows the odds ratio estimates of the main and interaction effects from the final logistic regression. Stronger association between serum creatinine and AKI has been observed in men (OR=1.014 per mg/dL increase) than in women (OR=1.003 per 100 unit increase), which is consistent with published reports. Biologically, men

have higher muscle mass and thus higher baseline serum creatinine, and KDIGO's absolute/relative sCr thresholds may differentially trigger AKI detection across sexes.

Liver disease history was strong risk in women (OR = 3.71) and in men (OR=3.04) with modest attenuation in men (interaction ROR = 0.82).

Sex modified several other predictors, with generally modest effects. In men, higher bicarbonate (OR=0.91 per mmol/L) strengthen the original protection effect, compared with women (OR=0.93; interaction ROR=0.98). Hemoglobin modestly attenuated the protective effect of higher hemoglobin in men (OR=0.93 per g/dL) than in women (OR=0.86 per g/dL). Mechanical ventilation in prior 30 days had a higher risk for AKI in men (OR=3.25) than in women (OR=2.92). Infection is strongly associated with AKI in both women (OR=3.48) and men (OR=3.36). The unadjusted distributions of these predictors by sex and AKI status in both MIMIC-III and MIMIC-IV are shown in Figure 3, which are consistent with adjusted effects except that infection history was barely captured in MIMIV-III.

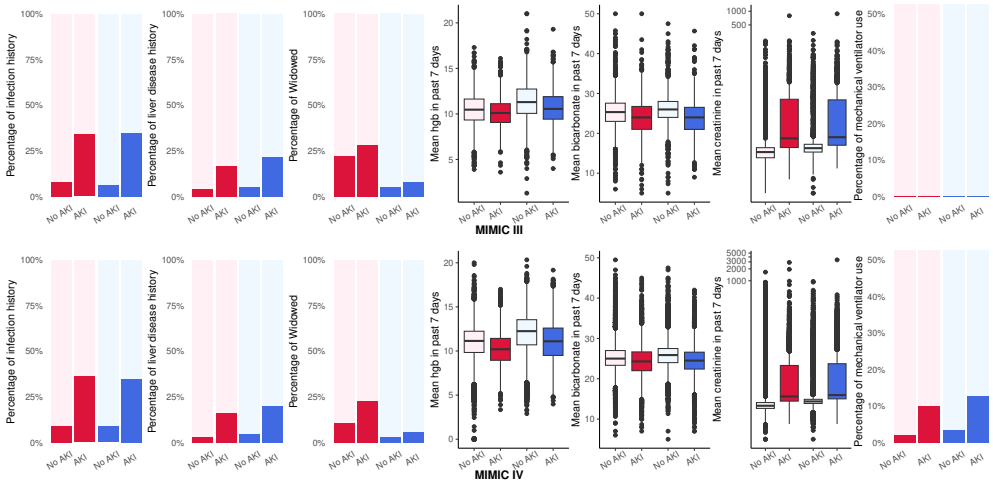

Fig. 3 Distributions of biomarkers in sex-specific AKI and no AKI groups. Red and blue are for females and males, respectively.

Based on the H-statistics from the random forest model C2, the strongest sex–predictor interactions include age, private insurance, single marital status, and Black race, suggesting that aging, access to care, social context, and race intersected with sex to shape risk. Cardiovascular comorbidities, including hypertension, coronary artery disease, and cerebrovascular disease, also ranked highly, implying sex-specific vulnerability along vascular pathways. Among biomarkers, the 7-day mean creatinine showed the leading interaction with sex, consistent with biologic and diagnostic differences. Because H-statistics capture nonlinear, non-additive structure that standard logistic regression can miss, these signals complement our regression findings. Further

Table 3 Adjusted Effects for AKI in Study Cohort (sex–age–specific mean imputation)

Variable (units)	Odds Ratio (95% CI)	p-value
Demographic		
Age (per year)	1.012 (1.011, 1.013)	$< 2.2 \times 10^{-16}$
Male (vs. female)	1.125 (0.776, 1.629)	0.535
Other Race (vs. White)	0.954 (0.890, 1.023)	0.186
Single (vs. Married)	1.185 (1.142, 1.230)	$< 2.2 \times 10^{-16}$
Widowed (vs. Married)	1.210 (1.140, 1.285)	3.9×10^{-10}
Unknown (vs. Married)	1.040 (0.959, 1.127)	0.344
Private Insurance (vs. Medicare)	0.905 (0.871, 0.940)	3.3×10^{-7}
Unknown Insurance (vs. Medicare)	0.784 (0.724, 0.848)	1.6×10^{-9}
Medical History		
Obesity	1.342 (1.277, 1.411)	$< 2.2 \times 10^{-16}$
Hypertension	1.432 (1.384, 1.483)	$< 2.2 \times 10^{-16}$
Diabetes	1.433 (1.382, 1.487)	$< 2.2 \times 10^{-16}$
Coronary artery disease	1.185 (1.138, 1.233)	1.0×10^{-16}
COPD	1.237 (1.183, 1.294)	$< 2.2 \times 10^{-16}$
Cancer	1.688 (1.619, 1.760)	$< 2.2 \times 10^{-16}$
Heart failure	2.396 (2.291, 2.505)	$< 2.2 \times 10^{-16}$
Infection	3.478 (3.307, 3.658)	$< 2.2 \times 10^{-16}$
Liver disease	3.709 (3.438, 4.000)	$< 2.2 \times 10^{-16}$
Cerebrovascular disease	0.958 (0.894, 1.026)	0.222
Biomarker		
WBC, 7-d mean (scaled $\times 100$)	1.794 (1.488, 2.163)	8.7×10^{-10}
Platelet, 7-d mean (scaled $\times 1000$)	0.700 (0.580, 0.845)	2.0×10^{-4}
Hemoglobin, 7-d mean (g/dL)	0.862 (0.849, 0.875)	$< 2.2 \times 10^{-16}$
Sodium, 7-d mean (mmol/L)	0.971 (0.968, 0.974)	$< 2.2 \times 10^{-16}$
Potassium, 7-d mean (mmol/L)	1.018 (1.006, 1.030)	0.002
Creatinine, 7-d mean (mg/dL)	1.003 (1.002, 1.004)	7.9×10^{-11}
BUN, 7-d mean (mg/dL)	1.024 (1.023, 1.025)	$< 2.2 \times 10^{-16}$
Bicarbonate, 7-d mean (mmol/L)	0.931 (0.923, 0.940)	$< 2.2 \times 10^{-16}$
Mechanical ventilation in prior 30 d	2.923 (2.675, 3.194)	$< 2.2 \times 10^{-16}$
Sex-Interaction		
Infection History $\times$ Male	0.967 (0.901, 1.037)	0.347
Liver disease $\times$ Male	0.820 (0.745, 0.903)	5.1×10^{-5}
Widowed $\times$ Male	0.838 (0.752, 0.933)	0.001
Bicarbonate $\times$ Male	0.982 (0.970, 0.995)	0.006
Hemoglobin $\times$ Male	1.079 (1.058, 1.100)	2.6×10^{-14}
Mechanical ventilation $\times$ Male	1.113 (0.993, 1.247)	0.065
Creatinine $\times$ Male	1.011 (1.008, 1.014)	1.2×10^{-16}

validation may help reveal clinically meaningful and robust effect and clarify actionable thresholds for sex-specific risk stratification.

Other main effects were directionally consistent and clinically coherent: higher WBC (OR=1.79) and potassium (OR=1.02 per mmol/L) increased AKI odds, whereas higher sodium (OR=0.97 per mmol/L), higher platelets (OR=0.70 per 1000-scale unit), and higher hemoglobin (above) were protective. Comorbidities, including hypertension (OR=1.43), diabetes (1.43), coronary artery disease (1.19), COPD (1.24),

cancer (1.69), heart failure (2.40), and obesity (1.34), were all independently associated with higher AKI risk, while cerebrovascular disease was not (OR=0.96; p=0.22). AKI association with demographic variables, including age, single/widowed marital status, Medicare insurance, are expected. The positive association with male was not significant, probably because observed sex differences in AKI risk were captured by sex–predictor interactions rather than by sex alone.

In the sensitivity analysis of admitted patients not due to automobile accidents (Table D), the AUC-ROC was slightly lower in MIMIC-IV but higher in MIMIC-III compared with the primary logistic model. The Widowed-male interaction was not selected in the new model; instead, a BUN-male interaction was selected, suggesting a stronger damaging effect of BUN in women (OR=1.36) than in men (OR=1.27).

The logistic model using MICE rather than filled means resulted in similar result except married status was significantly beneficial to men, but it was not observed in women (Table D).

3.4 Risk Factors for CKD

Among AKI patients in MIMIC-IV, men had a slightly higher cumulative incidence of CKD (12.3% vs. 12.0%) and developed it earlier (median 1.75 years; IQR 0.66–3.91 vs. 1.83 years; IQR 0.78–3.83) than women. In MIMIC-III, fewer than 2% of AKI patients developed CKD during follow-up, which did not allow for reliable sex comparisons.

CKD-free survival curves separated clearly by sex in both MIMIC-III and MIMIC-IV (Figure 4). By 5 years after AKI, 3-4% more men had developed CKD than women; by 10 years, the gap widened to about 5% in MIMIC-IV and 16% in MIMIC-III.

In the competing risk model of CKD after AKI (Table 3.4), CKD risk increased significantly with Black patinets (HR=1.15), divorced or windowed status(HR = 1.18), hypertension (HR = 1.27), diabetes (HR = 1.59), coronary artery disease (HR = 1.11), and heart failure (HR = 1.41). Patients with cancer, infection, or liver disease had lower CKD hazard (HRs = 0.63, 0.51, and 0.73, respectively), which was likely reflecting competing-risk mortality rather than true renoprotection. Among past 90-day laboratory markers, higher potassium increased CKD hazard (HR = 1.04 per mmol/L).

Sex modified several effects on CKD: the effect of cancer history on CKD hazard was attenuated in men compared with women (interaction HR = 1.26, p = 0.024); higher pH was protective in men (HR = 0.66 per unit) but not in women ; higher sodium increased CKD hazard in men (HR=1.01 per mmol/L) but reduced risk in women (HR= 0.99 per mmol/L); and creatinine had slightly stronger positive associations in men (interaction HR = 1.01 per mg/dL) but not in women.

For death (Table 3.4), mortality increased with age (HR = 1.02 per year), heart failure (HR = 1.29), COPD (HR=1.19), cancer (HR = 1.71), infection (HR = 2.59), liver disease (HR = 2.61), and cerebrovascular disease (HR = 1.62). Among 90-day laboratory markers, higher WBC (HR = 1.01 per unit), potassium (HR = 1.03 per mmol/L), and creatinine (HR = 1.004) were associated with slight increases in death hazard.

Unmarried status (single, divorced, widowed) showed HR $<$ 1 versus married, which reflects competing CKD event rather than true protective effect.

Sex modified the effect of heart failure on mortality: there was ignorable risk in men (HR = 1.00) but substantially stronger in women (HR = 1.29), with an interaction HR of 0.78 (p=4.8×10^{-5}).

Because deaths were less completely captured in MIMIC-IV than in MIMIC-III, the survival model derived from MIMIC-IV generalized poorly in MIMIC-III: AUC was close to 80% in MIMIC-IV but dropped to 65-75% in MIMIC-III. Conversely, MIMIC-III had many more early deaths and thus fewer CKD events, making CKD prediction less stable with AUC 55–65% (Figure E3).

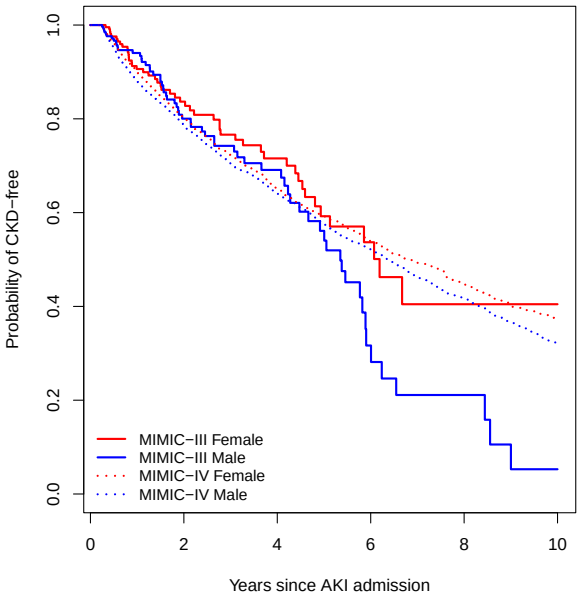

Fig. 4 Sex-specific post-AKI CKD onset curves. Red and blue are for females and males, respectively.

DeepSurv resulted in higher time-dependent AUC than the competing risk model, but it does not consider the competing nature of CKD an death. In DeepSurv, the most recent biomarker values dominated CKD risk prediction. Sex modified the effects of hypertension, cardiovascular disease, and potassium, which was not detected in the competing-risk model (Figure F5). For death, DeepSurv identified the same major risk factors as the competing-risk model but did not detect any sex modifiers.

4 Conclusion and Discussion

Our study result advances a sex-aware view of AKI and post-AKI CKD: risk differences were driven largely by sex–predictor interactions rather than sex alone. Demographics, insurance, comorbidity, and key blood markers yielded the strongest AKI discrimination, while CKD hazards rose with age, hypertension, diabetes, heart failure, and recent diuretics. The models will enable earlier identification of high-risk patients

and pinpointed predictors with sex-differential effects, informing sex-specific surveillance, medication stewardship, and risk communication. The stronger creatinine-AKI association in men, together with interactions for bicarbonate and hemoglobin, supports sex-specific thresholds and alerting. Collectively, these findings support targeted prevention and more precise post-AKI care pathways.

Our study had several limitations. First, the timing of AKI incidence was imprecise. Exact diagnosis timestamps were unavailable, so we anchored AKI at the hospital admission time and defined incident CKD as the first qualifying admission $\geq$90 days thereafter. This approximation might underestimate time to event and attenuate survival associations toward the null.

Second, our phenotyping relied on coded encounters within ICUs and emergency departments. Outpatient CKD or events treated elsewhere may be missed. The potential under-ascertainment and non-differential misclassification tend to bias results toward the null.

Third, missing data were substantial, especially for pre-admission biomarkers. We used a simple sex-age specific average filling in the primary analysis and observed similar estimates with multiple imputation under the missing-at-random assumption. Nonetheless, missingness may be informative (e.g., healthier patients are less frequently tested), violating MAR and potentially biasing effect estimates and calibration.

Fourth, several predictors were highly correlated and susceptible to confounding by indication or etiology. For example, white blood cell count tracks closely with infection, and trauma (e.g., automobile accidents) can correlate with both predictors and AKI risk. In sensitivity analyses that excluded automobile-related admissions, model performance and key associations were largely unchanged. The effects of those related variables might be underestimated due to positive correlations.

Finally, because MIMIC collected care data from predominantly critically ill patients, generalization of our study result might be limited and would require further validation before being extended to non-ICU or community settings.

There were only sparse longitudinal features prior to the AKI index time, thus no temporal data were encoded and inputted into the models. Thus some valuable within-subject correlation info might be lost, which guarantee the further investigations in a subgroups of patients.

Declarations

- Funding: Both RF and LL are supported by the DataWork!Prize Phase 1 award.
- Conflict of interest/Competing interests: Authors clared no conflict of interest.
- Code availability: https://github.com/noboundarystat/AKI-prediction-MIMIC
- Author contribution: Both authors contributed to data acquisition and analysis, result interpretation, and manuscript writing.

References

[1] J.A. Kellum, N. Lameire, Diagnosis, evaluation, and management of acute kidney injury: a kdigo summary (part 1). Kidney International **87**(1), 62–73 (2017)

[2] L.S. Chawla, P.L. Kimmel, R.L. Mehta, J.A. Kellum, Acute kidney injury and chronic kidney disease: an integrated approach. The Clinical Journal of the American Society of Nephrology **9**(12), 2062–2069 (2014)

[3] A.H. Zhang, Y. Huang, J. Zhao, L. Wang, H. Li, Gender differences in acute kidney injury: a systematic review and meta-analysis. Nephrology Dialysis Transplantation **35**(1), 23–31 (2020)

[4] C.Y. Hsu, C.E. McCulloch, C. Iribarren, J. Darbinian, A.S. Go, Race and gender differences in the risk of progression of chronic kidney disease. American Journal of Kidney Diseases **52**(2), 234–242 (2008)

[5] S. Bell, N. A., D. O., K. O., L. D., B. C., B. R., J. H., J. D., R. G., D. S., E. T., Gender differences in acute kidney injury. Clinical Journal of the American Society of Nephrology **11**(6), 988–998 (2016)

[6] N.H. Lameire, A. Bagga, D. Cruz, M. de Maeseneer, J. Delanghe, M. El Nahas, F. Finkelstein, S. Fishbane, G. Garcia-Garcia, C. Hsu, V. Jha, R. Kramar, G. Lonnemann, F. MacKenzie, A. Levin, Acute kidney injury: an increasing global concern. The Lancet **382**(9887), 170–179 (2013)

[7] E.A.J. Hoste, S.M. Bagshaw, S. Bell, M. Cerdan, A. Colloff, D. Cruz, V. D'Intini, Z. Endre, O. Gajic, J.L. Koyner, A. Lewington, C. Mammen, R.L. Mehta, D.J. O'Donoghue, P.M. Palevsky, J.W. Pickering, H. Ren, T. Tzeng, S.S. Waikar, A. Zarbock, Global epidemiology and outcomes of acute kidney injury. Nature Reviews Nephrology **14**(10), 607–625 (2018)

[8] S.C. Shih, H.J.Y. Ng, A.R. Choi, J.A. Hsu, Machine learning for predicting acute kidney injury: A systematic review. Kidney International **98**(3), 571–582 (2020)

[9] A.E.W. Johnson, T.J. Pollard, L. Shen, L.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L.A. Celi, R.G. Mark, Mimic-iii, a freely accessible critical care database. Scientific Data **3**, 160035 (2016)

[10] A.E. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L.A. Celi, R. Mark, MIMIC-IV, version 2.2. Scientific Data (2023). https://doi.org/10.1038/s41597-022-01899-x

[11] H. Jiang, W.H. Wu, J.L. Huang, C.H. Li, K.T. Chen, Using machine learning to predict acute kidney injury. Nephrology Dialysis Transplantation **36**(2), 286–293 (2021)

[12] N. Reimers, I. Gurevych, *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (Association for Computational Linguistics, Hong Kong, China, 2019), pp. 3982–3992. https://doi.org/10.18653/v1/D19-1410. URL https://aclanthology.org/D19-1410/

[13] N. Reimers, I. Gurevych, *Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation*, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Association for Computational Linguistics, Online, 2020), pp. 4512–4525. https://doi.org/10.18653/v1/2020.emnlp-main.365. URL https://aclanthology.org/2020.emnlp-main.365/

[14] L. Breiman, Random forests. Machine Learning **45**(1), 5–32 (2001)

[15] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp. 785–794 (2016)

[16] Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539. URL https://doi.org/10.1038/nature14539

[17] R. Wang, B. Fu, G. Fu, M. Wang, *Deep & Cross Network for Ad Click Predictions*, in *Proceedings of the 2017 Workshop on AdKDD & TargetAd* (ACM, 2017), pp. 12:1–12:7. https://doi.org/10.1145/3124749.3124754. URL https://arxiv.org/abs/1708.05123

[18] R. Wang, R. Shivanna, D. Cheng, S. Jain, D. Lin, L. Hong, E. Chi, *DCN V2: Improved Deep & Cross Network and Practical Lessons for Web-scale Learning to Rank Systems*, in *Proceedings of the Web Conference 2021 (WWW '21)* (ACM, 2021), pp. 1785–1797. https://doi.org/10.1145/3442381.3450078. URL https://arxiv.org/abs/2008.13535

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, *Attention Is All You Need*, in *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017), pp. 5998–6008. URL https://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf

[20] H. Zou, T. Hastie, Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**, 301–320 (2005)

[21] J. Katzman, U. Shaham, J. Bates, Y. Kluger, Deepsurv: Personalized treatment recommendation via a cox proportional hazards deep neural network. BMC Medical Research Methodology **18**(1), 24 (2018)

[22] P.J. Heagerty, T. Lumley, M.S. Pepe, Time-dependent roc curves for censored survival data and a diagnostic marker. Biometrics **56**(2), 337–344 (2000). https://doi.org/10.1111/j.0006-341X.2000.00337.x

[23] E. Graf, C. Schmoor, W. Sauerbrei, M. Schumacher, Assessment and comparison of prognostic classification schemes for survival data. Statistics in Medicine **18**(17-18), 2529–2545 (1999). https://doi.org/10.1002/(SICI)1097-0258(19990915/30)

18:17/18⟨2529::AID-SIM274⟩3.0.CO;2-5

[24] A. Fisher, C. Rudin, F. Dominici, All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research **20**(177), 1–81 (2019). URL http://jmlr.org/papers/v20/18-760.html

[25] T. Chen, C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, San Francisco, CA, USA, 2016), pp. 785–794. https://doi.org/10.1145/2939672.2939785. URL https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf

[26] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Liu, *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017). URL https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree

[27] S.M. Lundberg, S.I. Lee, *A Unified Approach to Interpreting Model Predictions*, in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (2017). URL https://arxiv.org/abs/1705.07874

[28] M. Sundararajan, A. Taly, Q. Yan, *Axiomatic Attribution for Deep Networks*, in *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)* (2017), pp. 3319–3328. URL http://proceedings.mlr.press/v70/sundararajan17a.html

[29] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.I. Lee, From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence **2**, 56–67 (2020). https://doi.org/10.1038/s42256-019-0138-9

[30] J.H. Friedman, B.E. Popescu, Predictive learning via rule ensembles. The Annals of Applied Statistics pp. 916–954 (2008)

[31] T.E. Raghunathan, J.M. Lepkowski, J. Van Hoewyk, P. Solenberger, A multivariate technique for multiply imputing missing values using a sequence of regression models. Survey Methodology **27**(1), 85–95 (2001). URL https://www150.statcan.gc.ca/n1/pub/12-001-x/2001001/article/5857-eng.pdf

[32] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate imputation by chained equations in r. Journal of Statistical Software **45**(3), 1–67 (2011). https://doi.org/10.18637/jss.v045.i03. URL https://www.jstatsoft.org/v45/i03/

Supplementary Information.

Appendix A Model Performance Metrics from all AKI prediction Models

Appendix B Feature Importance results from the top 30 features

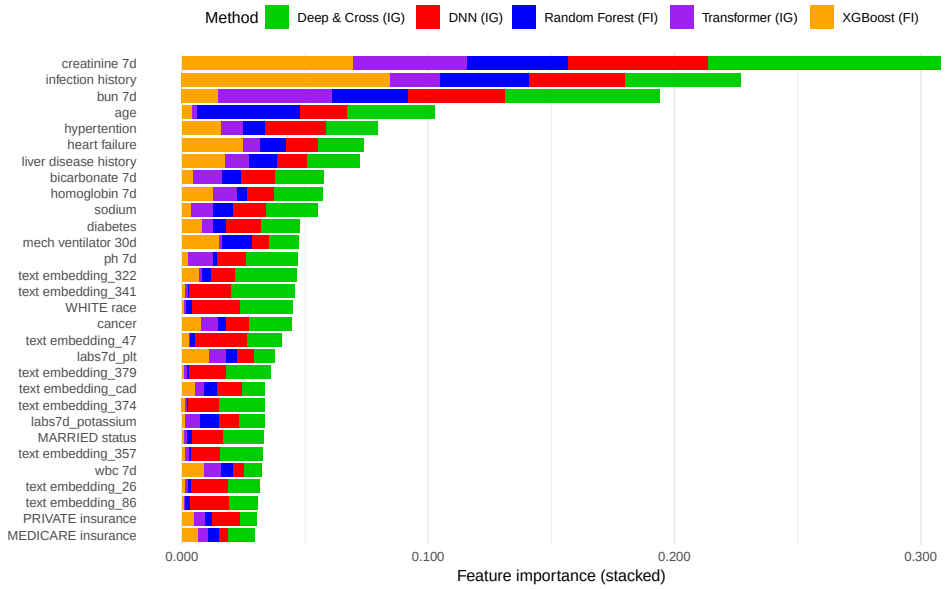

Fig. B1 Feature Importance of predictors in AKI prediction machine learning models. The top features are consistent with the predictors selected in logistic regression. in general, tree-based method and transformers are likely to prioritize explainable features, while DNN and DCN can pick many text embedding components.

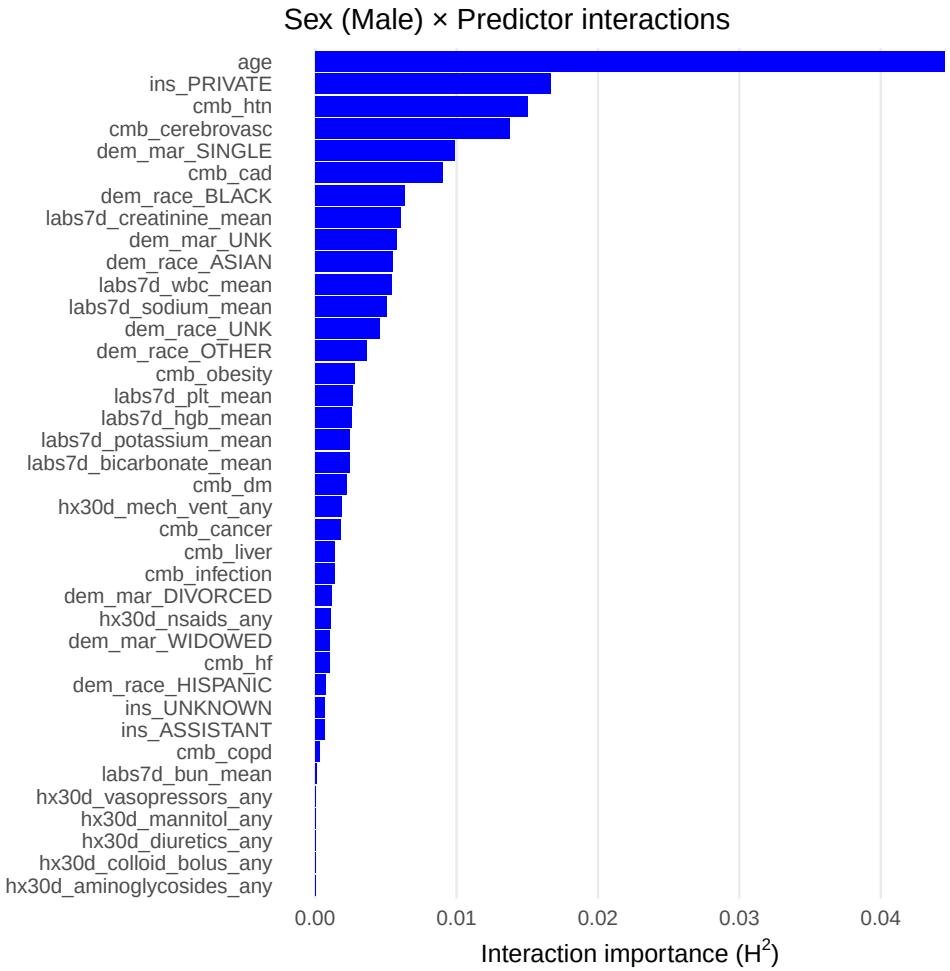

Fig. C2 Distribution of H-statistics for sex–biomarker interactions in the random forest model (MIMIC-IV)

Appendix C H^2 Statistic for Sex-Predictor Interactions in Random Forest Model

Appendix D Sensitivity Analysis in Subgroup and with alternative missing value imputation approach

Appendix E Evaluation of Competing Risk and DeepSurv Models for CKD and Death

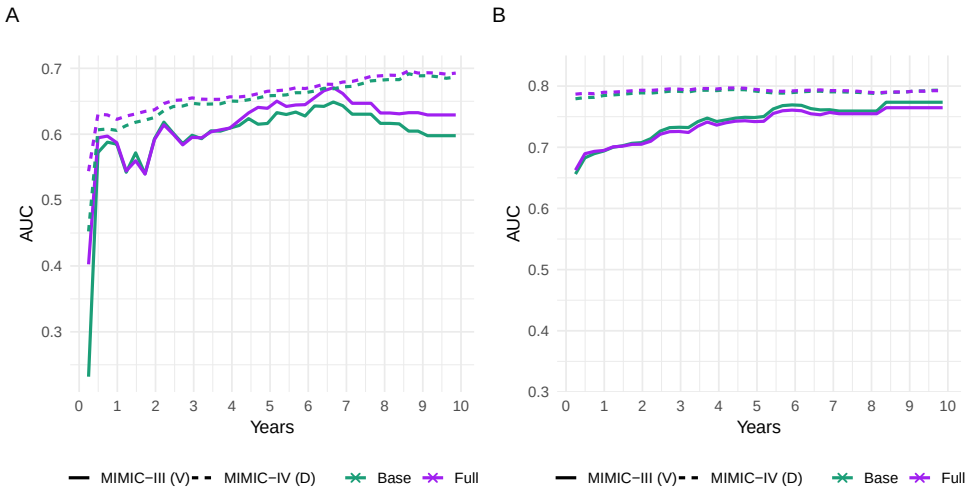

Fig. E3 Time-dependent AUC and Brier in MIMIC-IV (derivation) and MIMIC-III (validation)

Appendix F DeepSurv model fitting and feature importance

Appendix G Sensitivity Analysis for 60-day CKD

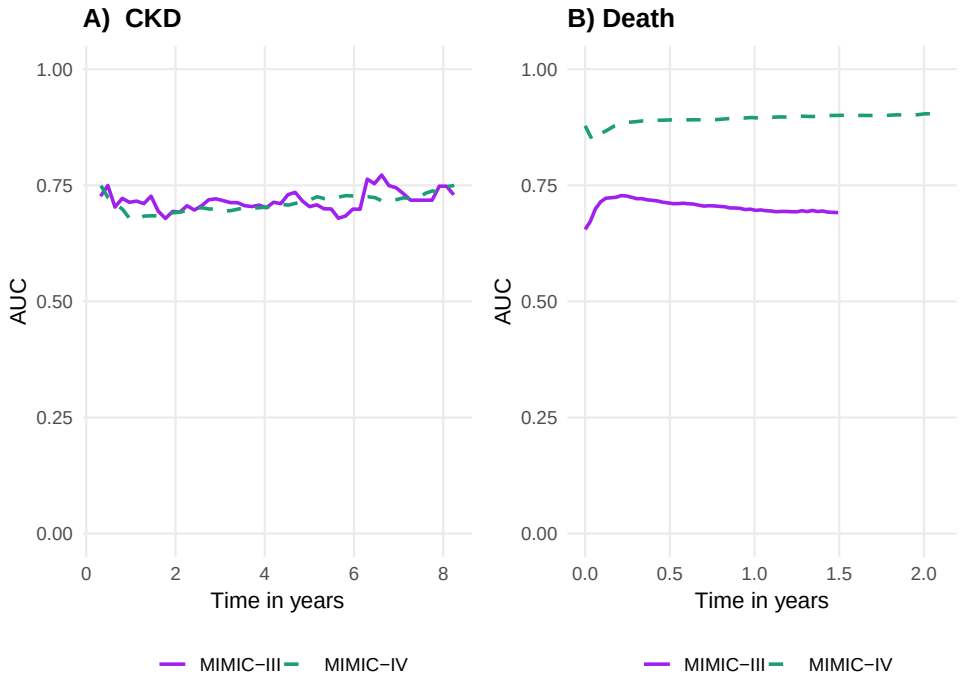

Fig. E4 Time-dependent AUC for MIMIC-III by DeepSurv

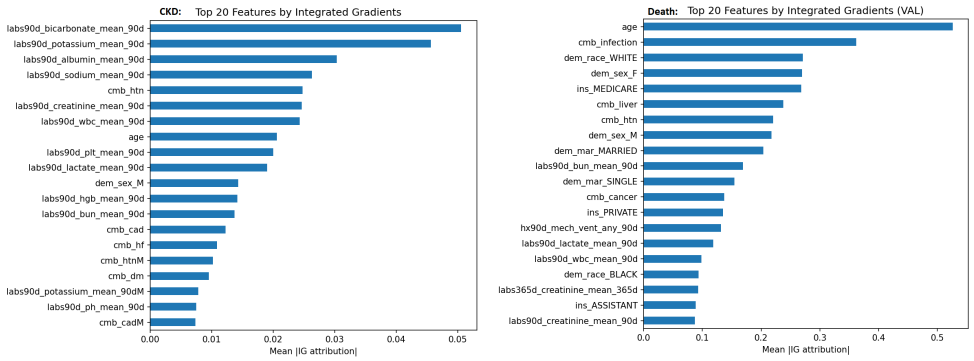

Fig. F5 The most important features for time to CKD prediction using Integrated Gradient in DeepSurv

Table 4 Adjusted Hazard Ratios for CKD and Death after AKI in Competing Risk Models

Outcome	Variable (units)	Hazard Ratio (95% CI)	p-value
CKD	**Demographic**		
	Age (per year)	1.003 (1.000, 1.006)	0.032
	Black race	1.154 (1.055, 1.261)	0.0016
	Divorced (vs. Married)	1.179 (1.039, 1.337)	0.011
	Widowed (vs. Married)	1.176 (1.058, 1.308)	0.0027
	Medical history		
	Hypertension	1.271 (1.174, 1.376)	3.6×10^{-9}
	Diabetes	1.594 (1.487, 1.710)	$< 2.2 \times 10^{-16}$
	Coronary artery disease	1.114 (1.031, 1.204)	0.0062
	Heart failure	1.412 (1.305, 1.527)	$< 2.2 \times 10^{-16}$
	Cancer	0.631 (0.537, 0.742)	2.5×10^{-8}
	Infection	0.505 (0.463, 0.550)	$< 2.2 \times 10^{-16}$
	Liver disease	0.726 (0.655, 0.805)	1.3×10^{-9}
	Biomarkers (90-day mean)		
	Sodium (mmol/L)	0.986 (0.974, 0.998)	0.022
	Potassium (mmol/L)	1.036 (1.008, 1.065)	0.012
	Sex interactions		
	Cancer × Male	1.260 (1.031, 1.542)	0.024
	pH (unitless) × Male	0.658 (0.560, 0.773)	3.3×10^{-7}
	Sodium (mmol/L) × Male	1.021 (1.013, 1.029)	1.0×10^{-7}
	Creatinine (mg/dL) × Male	1.011 (1.006, 1.016)	2.6×10^{-5}
Death	**Demographic**		
	Age (per year)	1.018 (1.015, 1.020)	$< 2.2 \times 10^{-16}$
	Single (vs. Married)	0.747 (0.693, 0.805)	1.9×10^{-14}
	Divorced (vs. Married)	0.726 (0.645, 0.817)	1.2×10^{-7}
	Widowed (vs. Married)	0.816 (0.738, 0.902)	7.0×10^{-5}
	Medical history		
	Heart failure	1.286 (1.172, 1.411)	1.1×10^{-7}
	COPD	1.186 (1.098, 1.281)	1.3×10^{-5}
	Cancer	1.711 (1.600, 1.830)	$< 2.2 \times 10^{-16}$
	Infection	2.592 (2.432, 2.762)	$< 2.2 \times 10^{-16}$
	Liver disease	2.614 (2.445, 2.796)	$< 2.2 \times 10^{-16}$
	Cerebrovascular disease	1.624 (1.463, 1.803)	$< 2.2 \times 10^{-16}$
	Biomarkers (90-day mean)		
	WBC ($\times 10^9$/L)	1.013 (1.010, 1.017)	5.0×10^{-12}
	Potassium (mmol/L)	1.029 (1.009, 1.050)	0.0053
	Creatinine (mg/dL)	1.004 (1.001, 1.008)	0.021
	Sex interactions		
	Heart failure × Male	0.777 (0.689, 0.878)	4.8×10^{-5}

Table A1 Model Evaluation Metrics

Method	AUC-ROC	Sensitivity/Recall	Specificity	Accuracy	Precision	F_1
MIMIC-IV testing set						
Logistic	0.851	0.790	0.751	0.756	0.324	0.459
Random Forest	0.909	0.864	0.814	0.821	0.414	0.560
XGboost	0.927	0.877	0.827	0.833	0.435	0.582
DNN	0.908	0.841	0.819	0.822	0.412	0.553
DCN	0.904	0.831	0.825	0.826	0.418	0.556
Transformer	0.916	0.845	0.832	0.834	0.431	0.571
MIMIC-III						
Logistic	0.774	0.681	0.735	0.724	0.390	0.496
Random Forest	0.777	0.767	0.646	0.670	0.351	0.481
XGboost	0.829	0.752	0.757	0.756	0.436	0.552
DNN	0.764	0.761	0.652	0.674	0.353	0.482
DCN	0.723	0.628	0.703	0.688	0.345	0.446
Transformer	0.739	0.651	0.702	0.692	0.353	0.457

Table D2 Adjusted Effects for AKI in Non-auto-accident Admissions with Sex-Age-Specific Mean Imputation for Missing Values

Variable (units)	Odds Ratio (95% CI)	p-value
Demographic		
Age (per year)	1.009 (1.008, 1.011)	$< 2.2 \times 10^{-16}$
Sex: Male (vs. female)	1.030 (0.651, 1.629)	0.900
Race: Other (vs. White)	1.051 (0.979, 1.129)	0.172
Marital status: Single (vs. married)	1.159 (1.109, 1.211)	5.3×10^{-11}
Marital status: Widowed (vs. married)	1.132 (1.061, 1.207)	1.7×10^{-4}
Insurance: Private (vs. Medicare)	0.870 (0.831, 0.911)	3.321×10^{-9}
Insurance: Unknown (vs. Medicare)	0.794 (0.723, 0.872)	1.364×10^{-6}
Medical History		
Obesity	1.375 (1.295, 1.458)	$< 2.2 \times 10^{-16}$
Hypertension	1.479 (1.418, 1.542)	$< 2.2 \times 10^{-16}$
Diabetes	1.463 (1.399, 1.529)	$< 2.2 \times 10^{-16}$
Coronary artery disease	1.203 (1.145, 1.264)	2.0×10^{-13}
Heart failure	2.600 (2.463, 2.744)	$< 2.2 \times 10^{-16}$
COPD	1.219 (1.150, 1.292)	2.9×10^{-11}
Cancer	1.537 (1.458, 1.619)	$< 2.2 \times 10^{-16}$
Infection	3.568 (3.363, 3.785)	$< 2.2 \times 10^{-16}$
Liver disease	3.584 (3.280, 3.915)	$< 2.2 \times 10^{-16}$
Cerebrovascular disease	0.939 (0.857, 1.029)	0.178
Biomarker		
WBC, 7-d mean (scaled $\times 100$)[a]	1.442 (1.134, 1.833)	0.003
Platelet, 7-d mean (1000×10^9/L)[a]	0.709 (0.562, 0.894)	0.004
Hemoglobin, 7-d mean (g/dL)	0.857 (0.841, 0.873)	$< 2.2 \times 10^{-16}$
Sodium, 7-d mean (mmol/L)	0.977 (0.973, 0.982)	$< 2.2 \times 10^{-16}$
Creatinine, 7-d mean (scaled $\times 100$)[a]	1.303 (1.179, 1.440)	2.235×10^{-7}
BUN, 7-d mean (per 10 mg/dL)[a]	1.363 (1.336, 1.390)	$< 2.2 \times 10^{-16}$
Potassium, 7-d mean (per 100 mmol/L)[a]	1.056 (0.202, 5.514)	0.949
Bicarbonate, 7-d mean (per 10 mmol/L)[a]	0.389 (0.348, 0.435)	$< 2.2 \times 10^{-16}$
Mechanical ventilation in prior 30 d	3.376 (3.002, 3.797)	$< 2.2 \times 10^{-16}$
Sex-Interaction		
Infection History × Male	0.926 (0.854, 1.005)	0.065
Liver disease × Male	0.742 (0.661, 0.832)	3.3×10^{-7}
Hemoglobin × Male	1.068 (1.043, 1.093)	4.1×10^{-8}
BUN (per 10) × Male	0.929 (0.906, 0.953)	1.2×10^{-8}
Bicarbonate (per 10) × Male	0.975 (0.833, 1.140)	0.749
Mechanical ventilation × Male	1.067 (0.919, 1.239)	0.393
Creatinine (per 100) × Male	3.589 (2.611, 4.933)	3.5×10^{-15}

[a] Variables labeled "scaled" are reported per the analysis coding (e.g., WBC ×100, platelets ×1000, creatinine ×100, BUN/bicarbonate per 10 units).

Table D3 Adjusted Effects for AKI in Non-auto-accident Admissions with Multiple Imputation for Missing Values

Variable (units)	Odds Ratio (95% CI)	p-value
Demographic		
Age (per year)	1.016 (1.015, 1.017)	$< 2.2 \times 10^{-16}$
Sex: Male (vs. female)	1.164 (0.826, 1.640)	0.384
Race: Black (vs. White)	1.552 (1.473, 1.636)	$< 2.2 \times 10^{-16}$
Race: Other (vs. White)	0.995 (0.926, 1.068)	0.883
Insurance: Assistant (vs. Medicare)	0.979 (0.923, 1.038)	0.476
Insurance: Private (vs. Medicare)	0.883 (0.838, 0.930)	2.6×10^{-6}
Insurance: Unknown (vs. Medicare)	0.784 (0.712, 0.863)	7.4×10^{-7}
Medical History		
Obesity	1.387 (1.308, 1.470)	$< 2.2 \times 10^{-16}$
Hypertension	1.454 (1.395, 1.515)	$< 2.2 \times 10^{-16}$
Diabetes	1.438 (1.377, 1.502)	$< 2.2 \times 10^{-16}$
Coronary artery disease	1.214 (1.156, 1.274)	5.3×10^{-15}
Heart failure	2.654 (2.517, 2.799)	$< 2.2 \times 10^{-16}$
COPD	1.244 (1.175, 1.318)	8.1×10^{-14}
Cancer	1.587 (1.508, 1.670)	$< 2.2 \times 10^{-16}$
Infection	3.575 (3.374, 3.788)	$< 2.2 \times 10^{-16}$
Liver disease	3.035 (2.866, 3.215)	$< 2.2 \times 10^{-16}$
Cerebrovascular disease	0.921 (0.841, 1.008)	0.074
Biomarker		
Platelet, 7-d mean (per 100×10^9/L)[a]	0.948 (0.931, 0.966)	2.2×10^{-8}
Hemoglobin, 7-d mean (g/dL)	0.905 (0.892, 0.918)	$< 2.2 \times 10^{-16}$
Sodium, 7-d mean (mmol/L)	0.822 (0.796, 0.848)	$< 2.2 \times 10^{-16}$
Creatinine, 7-d mean (scaled $\times 100$)[a]	0.853 (0.779, 0.933)	5.5×10^{-4}
BUN, 7-d mean (per 10 mg/dL)[a]	1.123 (1.113, 1.134)	$< 2.2 \times 10^{-16}$
Potassium, 7-d mean (per 100mmol/L)	0.868 (0.261, 2.886)	0.817
Bicarbonate, 7-d mean (per 10 mmol/L)[a]	0.514 (0.472, 0.559)	$< 2.2 \times 10^{-16}$
Mechanical ventilation in prior 30 day	2.377 (2.107, 2.682)	$< 2.2 \times 10^{-16}$
Sex-Interaction		
Married × Male	0.868 (0.825, 0.913)	4.8×10^{-8}
Infection History × Male	0.904 (0.835, 0.979)	0.013
BUN (per 10) × Male	0.996 (0.984, 1.008)	0.484
Hemoglobin × Male	1.032 (1.013, 1.052)	0.0007
Creatinine (per 100) × Male	1.472 (1.268, 1.710)	4.0×10^{-7}
Bicarbonate (per 10)× Male	1.054 (0.938, 1.183)	0.377
Mechanical ventilation × Male	1.094 (0.938, 1.275)	0.254

Platelets per 100×10^9/L; BUN and bicarbonate per 10 units; creatinine scaled by 100 (presented per scaled unit).

Table G4 Adjusted Hazard Ratios for 60d-CKD and Death after AKI in Competing Risk Models

Outcome	Variable (units)	Hazard Ratio (95% CI)	p-value
CKD	**Demographic**		
	Black race	1.120 (1.028, 1.221)	0.0093
	Divorced (vs. Married)	1.169 (1.036, 1.320)	0.012
	Widowed (vs. Married)	1.189 (1.078, 1.312)	5.5×10^{-4}
	Medical history		
	Hypertension	1.266 (1.175, 1.363)	5.6×10^{-10}
	Diabetes	1.607 (1.502, 1.719)	$< 2.2 \times 10^{-16}$
	Coronary artery disease	1.102 (1.023, 1.187)	0.011
	Heart failure	1.432 (1.328, 1.544)	$< 2.2 \times 10^{-16}$
	Cancer	0.669 (0.573, 0.780)	3.2×10^{-7}
	Infection	0.532 (0.490, 0.577)	$< 2.2 \times 10^{-16}$
	Liver disease	0.762 (0.691, 0.840)	5.1×10^{-8}
	Biomarkers (90-day mean)		
	Hemoglobin (g/dL)	1.058 (1.022, 1.095)	1.2×10^{-3}
	Sodium (mmol/L)	0.987 (0.977, 0.998)	0.018
	BUN (mg/dL)	1.003 (1.001, 1.005)	3.3×10^{-3}
	Potassium (mmol/L)	1.035 (1.007, 1.065)	0.015
	Sex interactions		
	Cancer × Male	1.224 (1.010, 1.483)	0.039
	pH (unitless) × Male	0.720 (0.611, 0.848)	8.9×10^{-5}
	Sodium (mmol/L) × Male	1.017 (1.009, 1.025)	3.6×10^{-5}
Death	**Demographic**		
	Age (per year)	1.017 (1.015, 1.020)	$< 2.2 \times 10^{-16}$
	Single (vs. Married)	0.754 (0.700, 0.813)	1.4×10^{-13}
	Divorced (vs. Married)	0.733 (0.651, 0.825)	2.7×10^{-7}
	Widowed (vs. Married)	0.814 (0.736, 0.900)	5.5×10^{-5}
	Medical history		
	Heart failure	1.283 (1.170, 1.408)	1.3×10^{-7}
	COPD	1.195 (1.107, 1.291)	5.3×10^{-6}
	Cancer	1.690 (1.580, 1.807)	$< 2.2 \times 10^{-16}$
	Infection	2.551 (2.393, 2.718)	$< 2.2 \times 10^{-16}$
	Liver disease	2.590 (2.423, 2.770)	$< 2.2 \times 10^{-16}$
	Cerebrovascular disease	1.606 (1.447, 1.783)	$< 2.2 \times 10^{-16}$
	Biomarkers (90-day mean)		
	WBC ($\times 10^9$/L)	1.014 (1.010, 1.018)	1.8×10^{-12}
	Potassium (mmol/L)	1.033 (1.012, 1.055)	0.0018
	Sex interactions		
	Heart failure × Male	0.769 (0.681, 0.868)	2.1×10^{-5}