Original Article

Characteristics and clinical subtypes of cancer patients in the intensive care unit: a retrospective observational study for two large databases

Shaowei Gao[#], Yaqing Wang[#], Lu Yang, Zhongxing Wang, Wenqi Huang

Department of Anesthesia, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Contributions: (I) Conception and design: W Huang, Z Wang, S Gao; (II) Administrative support: W Huang, Z Wang; (III) Provision of study materials or patients: S Gao; (IV) Collection and assembly of data: S Gao, Y Wang; (V) Data analysis and interpretation: S Gao, Y Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Wenqi Huang, MD; Zhongxing Wang, MD. Department of Anesthesia, the First Affiliated Hospital of Sun Yat-sen University, 58 Zhongshan 2nd Road, Guangzhou 510080, China. Email: huangwq@mail.sysu.edu.cn; wzhxing@mail.sysu.edu.cn.

Background: Previous studies have reported very different mortality rates among cancer patients in the intensive care unit (ICU), implying different clinical subtypes. We aimed to reveal the clinical subtypes and demonstrate the importance of segregating the patients in clinical research, and to report the ICD-level mortality of cancer patients in the ICU.

Methods: Two ICU databases (MIMIC-III and eICU) were utilized to identify cancer patients. Mortality based on ICD-level diagnoses were calculated, and K-means clustering was used to identify different clinical subtypes in the MIMIC database. Clinical characteristics and outcomes were compared among subtypes, and the calibration of SAPS II and APACHE IV among different subtypes was evaluated.

Results: In total, 6,505 (13.8%) cancer patients of the MIMIC database and 7,351 (4.9%) ones in eICU database, were enrolled in the study. Metastasis involving pleura, metastasis involving the liver, and acute myeloid leukemia were in the top 5 diagnoses with the highest mortality in both databases. Clinical subtypes identified by K-means clustering were closely associated with admission type (elective or emergency) and clinical service provider (surgical or medical). In a four-cluster pattern, nearly all patients in the first cluster were elective admissions (99.1%), whereas in the rest of the clusters, most were emergency admissions (93.7%). Most surgical patients were in the 1+2 clusters (92.0%) and most medical patients were in the 3+4 clusters (93.5%). Most characteristics and outcomes as well as the calibration of SAPS II and APACHE IV scoring systems were significantly different among clinical subtypes.

Conclusions: Different clinical subtypes can be well identified by admission type and clinical service provider among ICU patients with cancer. Caution should be exercised when considering these patients as a whole population both in clinical practice and research. Moreover, APACHE IV has better calibration than SAPS II for cancer patients at low risk of mortality in the ICU.

Keywords: Cancer mortality; intensive care unit (ICU); clinical subtypes; scoring system; K-mean

Submitted Jun 11, 2020. Accepted for publication Oct 10, 2020.

doi: 10.21037/atm-20-4634

View this article at: http://dx.doi.org/10.21037/atm-20-4634

Introduction

Cancer is a common comorbidity in patients admitted to the intensive care unit (ICU). Approximately one in seven ICU patients have cancer comorbidity (1,2). Previous studies have explored the characteristics and outcomes of critically ill cancer patients and reported highly varied mortality for this population (3-5). These mortality-related variations indicate the difficulty of prognostication in critically ill cancer patients (6). Some exceptional research has specified criteria to select these patients such as postoperative admission after elective surgery (4,7). These studies always had similar outcomes, whereas others had quite different outcomes. The disparity in the mortality rates of ICU patients with cancer may imply some clinical subtypes that are so different in clinical features and prognoses that they cannot be regarded as a whole population, both in clinical practice and research. Therefore, to identify these subtypes may help explain the abovementioned disparity in mortality rates and facilitate prognostication in this specific population.

Furthermore, previous studies only revealed varied mortality among different types of cancer at organ level (8,9). Cancers at different locations within the same organ have distinct prognoses, such as for hilar versus peripheral lung cancer (10). Therefore, more detailed evidence, particularly on mortality at International Classification of Diseases (ICD) level, is needed.

We conducted this study to explore the characteristics and outcomes, particularly the ICD-level mortality rate, of patients with cancer in the ICU. Moreover, we introduced an unsupervised machine learning algorithm to identify different patient subpopulations through the analysis of cancer patients from two public ICU databases derived from large medical centers. We also evaluated the calibration of SAPS II and APACHE IV in this population. We present the following article in accordance with the STROBE reporting checklist (available at http://dx.doi.org/10.21037/atm-20-4634).

Methods

Data source

We extracted data from two large public databases: the Medical Information Mart for Intensive Care (MIMIC-III) Database (https://mimic.physionet.org) and eICU Collaborative Research Database (https://eicu-crd.mit.edu) (11,12). The MIMIC database contains medical records of more than 50,000 distinct admissions in the critical care units of the Beth Israel Deaconess Medical Center that occurred between 2001 and 2012, whereas the eICU database collected medical records from over 200,000 admissions from multiple centers in 2014 and 2015.

Extraction criteria

The inclusion criteria were: (I) having diagnosis code of neoplasm (ICD-9 codes: 140–209) at any discharge diagnosis position; (II) the length of ICU stay was more than 4 hours; and (III) age between 16 and 89. We excluded ICU readmissions at the same hospital; in case of inter-ICU shift, only the first ICU stay was considered.

Data extraction and definition of variables

For both databases, the ICD-9 codes, along with their priority in each patient's diagnoses (a sequence in the MIMIC database, but classified as "primary," "main," "other" in the eICU database), were extracted to identify specific cancer types.

In the MIMIC database, we extracted data on demographics (age, sex, type of admission, and type of care unit), clinical service provider (surgical or medical), main comorbidities, vital signs, laboratory parameters reflecting illness severity, and organ-supporting treatment (mechanical ventilation, vasopressor, and renal replacement treatment) for analysis.

The clinical service referred to the service that a patient was admitted under. The main comorbidities were identified as previously described (13), and classified as congestive heart failure (CHF), atrial fibrillation (AFIB), liver disease, chronic obstructive pulmonary disease (COPD), chronic renal disease (CRD), coronary artery disease (CAD), stroke according to ICD-9 codes. The three abovementioned organ-supporting treatments were defined as binomial variables: whether they were used during the first 24 hours. Severity scoring systems, including the Simplified Acute Physiology Score II (SAPS II), Sequential Organ Failure Assessment (SOFA), and the Elixhauser Comorbidity Score, were calculated based on the respective published papers (14-16), with the code posted in the Github repository of the MIMIC database (17).

For the eICU database, the SAPS II score was calculated similarly as for the MIMIC database. However, the APACHE IV system was integrated in itself. Therefore, we extracted the APACHE IV score, predictive hospital

mortality based on APACHE IV, the calculated SAPS II scores, and actual hospital mortality from the eICU database.

Outcomes

The main outcome in the study was hospital mortality; length of ICU stay was demonstrated during the comparison of clinical subtypes.

Statistical analysis

The morbidity and mortality of cancer patients in the ICU were calculated within ICD-9 level and visualized by a histogram. For each type of cancer, the diagnostic priority was demonstrated by a heatmap. As all patient diagnoses were sequenced in the MIMIC database, we normalized the diagnostic space into 10 segments (0%~, 10%~, ...) and assigned those diagnoses into segments by their priority—that is, the primary diagnosis was at the 0%-segment and the last diagnosis was at the 90%-segment. Diagnostic priority was normalized by the Eq. [1]:

$$p_N = \begin{cases} 0\%(p_{max}=1) \\ \frac{p-1}{p_{max}-1}\times 100\%(p_{max}>1) \end{cases} \quad [1]$$

Where p_N refers to the normalized priority of the diagnosis, p_{max} refers to the total number of the patient's diagnoses, and p refers to the priority of the diagnosis before they were normalized. For each type of cancer, the sum of diagnoses in each segment was calculated. When visualized on the heatmap, the sum of diagnoses of each segment was normalized by the maximal value of all segments, as shown in Eq. [2]:

$$S_N = \frac{S}{S_{max}} \quad [2]$$

Where S_N refers to the normalized sum of diagnoses in each segment, S refers to the actual sum of diagnoses in each segment, and S_{max} refers to the sum of diagnoses in the segment with the largest number of diagnoses.

For the eICU database, we repeated the process above, but retained the three original diagnostic priorities ("primary," "main," and "other") instead of ten segments.

We used an unsupervised machine-learning method, K-means clustering, to cluster the patient population into subtypes (18). Briefly, K-means clustering aims to partition n observations into k clusters according to the nearest Euclidean distance to the centroid of the belonging cluster (18). The centroid of each cluster, which is randomized at first, iteratively updates by calculating the average point of each cluster. Finally, observations with similar patterns are partitioned into the same cluster. Most of the abovementioned variables were used in the K-means model except for the severity-scoring systems and outcomes. Unlike variables calculated in the severity-scoring systems, we used the mean values of vital signs and laboratory parameters in the first 24 hours of ICU admission rather than the worst values. Missing values of variables fed into the K-means model were imputed with a logistic regression algorithm that predicted the missing value by other variables fed into the K-means model.

An "elbow plot" was drawn to help determine the best number of clusters. If failed, we would select it from different total numbers [2–4] of clusters (19). The results of the K-means clustering were visualized by a heatmap that demonstrated the relative values of all variables. Then, with an algorithm named principal component analysis (PCA) (20), all variables of each patient were compressed into two components (PCA1 and PCA2), whereby patients were visualized on a scatterplot with two dimensions that indicated the relationship between patients and clusters.

As the MIMIC database provided more detailed clinical data, we used the K-means method only for MIMIC database, and considered the eICU as the validation database. After clustering, variables were compared among different clusters, which implied different clinical subtypes. Variables that were closely related to the clusters were used as instrumentation variables to identify different clinical subtypes. We evaluated the interactions of clinical subtype, the severity scoring system, and outcomes by using the standard mortality ratio (SMR), calibration graph, and Hosmer-Lemeshow C statistic (21). The SMR was calculated by a ratio of the observed mortality divided by predicted mortality; 95% confidence intervals were calculated by Vandenbroucke method (22).

In this study, continuous and categorical data were summarized as mean (SD) and number (percentage) respectively. One-way Analysis of variance (ANOVA) and chi-square test were used for statistical inferences of continuous and categorical data. A P value less than 0.05 was considered indicative of statistically significant differences. Databases were created locally with Postgresql (version 10.0) and accessed by structured query language integrated into a Python 3.7 notebook, on which run all the data-analysis codes used in the study. The notebook was

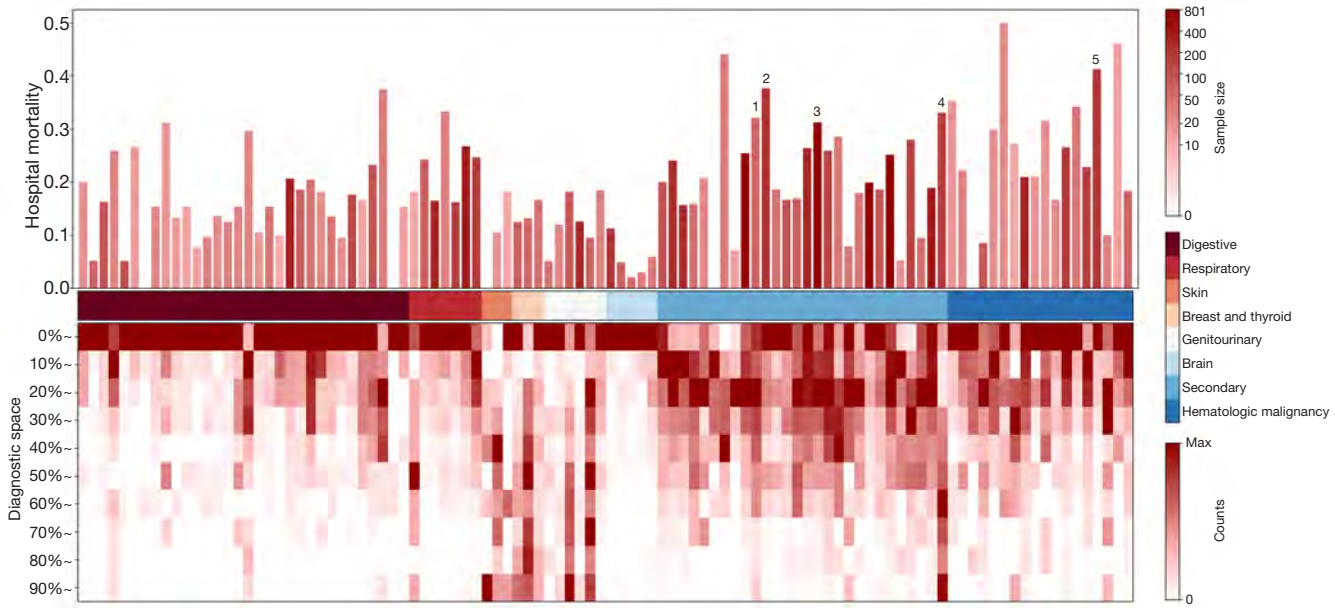

Figure 1 The mortality and distribution of positions of ICD-level diagnoses of the MIMIC database. In the top part of the figure, the depth of red color corresponds to the number of patients with a specific cancer. In the middle, the color blocks refer to the manually classified cancers according to the location or organ system. In the bottom part of the figure is a heatmap. The diagnostic space of each patient was divided into 10 segments, and the color in each segment refers to the frequency of the diagnosis occurring in that segment. The annotations were: 1, secondary malignant neoplasm of the mediastinum; 2, secondary malignant neoplasm of the pleura; 3, malignant neoplasm of the liver, secondary; 4, other malignant neoplasm without specification of the site; 5, acute myeloid leukemia, without details of having achieved remission. ICD, International Statistical Classification of Diseases and Related Health Problems; MIMIC, Medical Information Mart for Intensive Care.

uploaded on the Github for maintaining reproducibility (https://github.com/gscfwid/Cancer_ICU).

Results

According to extractive criteria, we extracted data on 6,505 (13.8%) patients with cancer from 46,998 patients with non-specific diagnosis from the MIMIC database. Among these cancer patients, cancer-related diagnostic codes were counted 11,174 times. While in the eICU database, up to 7,351 (4.9%) patients with cancer were extracted from 148,532 patients with nonspecific diagnoses, and the cancer-related diagnosis was counted 7,893 times. The most common cancer types based on the ICD-9 system were metastatic cancer of bone/bone marrow in MIMIC and lung cancer in the eICU database. Moreover, the diagnosis with the highest mortality (for cancer type with more than 10 patients) was mycosis fungoides in the MIMIC database, and metastatic adenocarcinoma of unknown primary in the eICU database. The top 10 most common and most fatal cancer types in the ICD level were summarized in the Tables S1,S2.

Mortality and diagnostic priority of different cancer types that had more than 10 victims are shown in *Figure 1*. The top part of *Figure 1* depicts the mortality rates of all types of cancer. The top 5 cancer types with the highest mortality that had at least 50 victims were annotated. The middle of the figure lists manually classified cancer types according to the organ or system, and this manual classification is listed respectively in Tables S3,S4 in the MIMIC and eICU databases, respectively. The priorities for each diagnosis were illustrated in the bottom part of *Figure 1*. We analyzed the eICU database in a similar way (Figure S1). Generally, for both databases, ICU patients with metastatic cancer, hematologic malignancy had higher mortality rates than others; they were less likely to be the primary diagnosis, except in metastatic brain tumor. Moreover, in the MIMIC database, most non-metastatic solid cancers held the first position at diagnosis.

As the elbow plot hardly provided the best cluster

number (Figure S2), we conducted a spectrum of K-means clustering with different cluster numbers [2–4] for the MIMIC database, as illustrated in *Figure 2A* and Figures S3,S4. The boundary of all 4 clusters were relatively clear (*Figure 2B*). Variables that were mostly associated with different clusters were admission type (elective or emergency) and clinical service provider (surgical or medical). We noticed that nearly all patients in the first cluster from the clustering of K=4 (*Figure 2*) were elective admissions (99.1%), whereas in the rest of the clusters, most were emergency admissions (93.7%). Most surgical patients were in the 1+2 clusters (92.0%) and most medical patients were in the 3+4 clusters (93.5%). Detailed disparities of the characteristics and outcomes from all four clusters are summarized in *Table 1*.

When comparing the clusters with different K numbers (*Figure 2A*, Figures S3 and S4), an evolutionary process could be observed. When K=2, it reflected the differences between clinical service providers (surgical or medical). When K=3, a new cluster was generated and featured worse cardiac and renal function. Finally, in the clustering of K=4, the cluster that featured with surgical service was divided into elective and emergency admissions.

Interestingly, although we did not use any outcome data or severity scoring data in the K-means model, the clusters were extremely associated with outcomes. The mortality rates of the four clusters were 4.0%, 11.8%, 20.4%, and 46.5%, which showed similar discrimination to the SAPS II system (4.8%, 8.5%, 20.1%, and 50.8%), as illustrated in *Figure 2C*. Here, we sorted all patients according to the SAPS II score in ascending order and divided them into four groups which had the same sizes with the K-means clusters. The mortality and SAPS II score were compared between these two systems. Except for the fourth cluster, we noticed similar SAPS II scores but distinct mortality among the clusters. Especially in the second and third clusters, the median and interquartile range of the SAPS II score was very close {37 [30–45] *vs.* 38.5 [31–47]}, but the mortality of the second cluster was nearly half that of the third cluster (11.8% *vs.* 20.3%; *Figure 2C*).

Furthermore, we validated the abovementioned findings in the eICU database. Because the clinical subtypes of cancer patients in the ICU were actually based on the clinical service provider, we classified patients of the eICU database into surgical and non-surgical groups based on the ICU admission resource: patients from the operation room were identified as belonging to the surgical group; otherwise, they were assigned to the non-surgical group, corresponding to clusters 1+2 and 3+4 in the MIMIC database. The SMR, Hosmer–Lemeshow C statistic, and calibration graph were conducted on the basis of different scoring systems (*Table 2* and *Figure 3*), and the corresponding number of patients in each risk decile were plotted in Figure S5. The calibration of SAPS II for both databases were poor (C=628.07, P<0.001 for the MIMIC database; C=827.64, P<0.001 for the eICU database; as shown in *Table 2* and *Figure 3A,3B*). However, the calibration of APACHE IV in the eICU database was relatively better, despite a P value less than 0.05 (C=87.84, P<0.001; as shown in *Table 2*, *Figure 3C,3D*). Furthermore, the SMRs of the surgical groups (cluster 1+2 in the MIMIC database) were lower than those in the non-surgical groups (cluster 3+4 in the MIMIC database) for all comparisons (*Table 2* and *Figure 3*).

The number of patients in each risk decile of the APACHE IV were imbalanced: numbers in high-risk deciles were fewer (Figure S5). Therefore, we compared the actual and predicted mortality for lower risk subgroups (*Figure 4*). The population of each database was divided into segments according to different SPAS II or APACHE IV scores, and the actual mortality, with 95% CIs, of the surgical and non-surgical group in each segment were compared, along with the predicted mortality by scoring systems.

For both databases, the mortality of the surgical group (cluster 1+2 in MIMIC database) was significantly lower than that of the non-surgical group (cluster 3+4 in MIMIC database; *Figure 4A,4B*). However, as the mortality predicted by the APACHE IV was calculated by the APACHE score and admission information, including diagnoses and admission information, surgical and non-surgical groups had different predicted mortality (*Figure 4C*). As shown in *Figure 4C*, the actual mortality was close to the mortality predicted by APACHE IV in these lower risk segments (the mortality of those segments was less than 0.25).

Discussion

The present study showed that, according to the ICD-level mortality (*Figure 1*, Figure S1), secondary and hematologic malignancies have worse outcomes. Moreover, among the top 5 diagnoses that had the highest mortality rates in the two databases, three of the diagnoses were the same: (I) metastasis involving pleura; (II) metastasis involving the liver; and (III) acute myeloid leukemia. This is consistent with the mainstream opinion that metastatic cancer and hematologic malignancy are independent prognostic

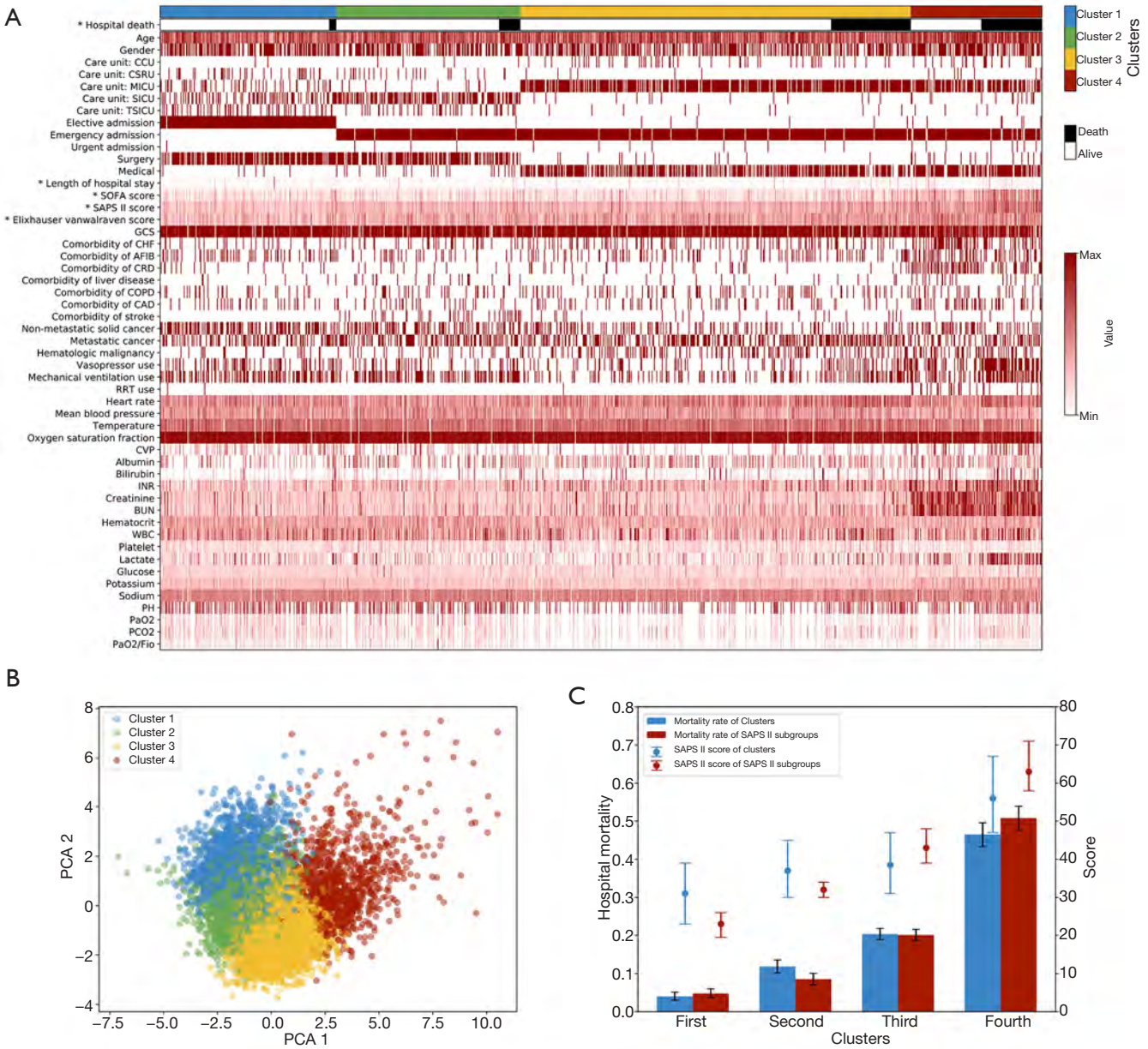

Figure 2 K-means clustering revealed four clinical subtypes (cluster 1-4) (A,B) and their associations with hospital mortality as compared with the SAPS II system (C). In Subplot A, the depth of red color in the heatmap referred to the relative value of the variables, and their names are annotated on the left. Variables marked by asterisk were not entered into the K-means model. The Subplot B refers to the visualization of Principal Component Analysis (PCA), and PCA1 and PCA2 are the two largest components. Each point represents an individual patient. The mortality of the four clusters are illustrated in Subplot C, compared with the four groups divided by the SAPS II score (the four SAPS II groups had the same sizes as those four clusters and their SAPS scores were sorted in an ascending order). CCU, cardiac care unit; CSRU, Cardiac surgery recovery unit; MICU, medical intensive care unit; SICU, surgery intensive care unit; TSICU, trauma/surgical intensive care unit; SOFA, sequential organ failure assessment; SAPSII, Simplified Acute Physiology Score II; GCS, Glasgow Coma Scale; CHF, chronic heart failure; AFIB, atrial fibrillation; CRD, chronic renal disease; COPD, chronic obstructive pulmonary disease; CAD, coronary artery disease; RRT, renal replacement treatment; CVP, central venal pressure; INR, international normalized ratio; BUN, blood urea nitrogen; WBC, white blood cells; PaO2, artery partial pressure of oxygen; PCO2, partial pressure of carbon dioxide; FiO2, fraction of inspired oxygen.

Table 1 Comparison of the characteristics and outcomes of the four clusters from the K-means clustering

	Missing	Cluster 1	Cluster 2	Cluster 3	Cluster 4	P value
n		1,307	1,379	2,785	1,034	
Age[a]	0	64.3 (12.4)	64.5 (13.2)	64.3 (13.9)	69.9 (12.7)	<0.001
Gender (M)[a]	0	796 (60.9)	831 (60.3)	1,528 (54.9)	700 (67.7)	<0.001
Care unit[b]	0					<0.001
CCU		15 (1.1)	34 (2.5)	327 (11.7)	119 (11.5)	
CSRU		235 (18.0)	168 (12.2)	16 (0.6)	23 (2.2)	
MICU		239 (18.3)	67 (4.9)	2,370 (85.1)	785 (75.9)	
SICU		530 (40.6)	825 (59.8)	42 (1.5)	81 (7.8)	
TSICU		288 (22.0)	285 (20.7)	30 (1.1)	26 (2.5)	
Admission type[b]	0					<0.001
Elective		1,258 (96.3)	1 (0.1)	56 (2.0)	59 (5.7)	
Emergency		1 (0.1)	1,367 (99.1)	2,674 (96.0)	951 (92.0)	
Urgent		48 (3.7)	11 (0.8)	55 (2.0)	24 (2.3)	
Service provider[b]	0					<0.001
Surgery		1,043 (79.8)	900 (65.3)	95 (3.4)	77 (7.4)	
Medical		35 (2.7)	196 (14.2)	1,957 (70.3)	790 (76.4)	
Unclear		229 (17.5)	283 (20.5)	733 (26.3)	167 (16.1)	
SOFA score[a]	0	3.4 (2.4)	3.3 (2.5)	3.6 (2.6)	8.0 (3.8)	<0.001
SAPS II score[a]	0	31.6 (12.5)	38.5 (12.0)	39.1 (11.3)	57.6 (15.5)	<0.001
Elixhauser Vanwalraven score[a]	10	8.8 (7.5)	12.3 (8.1)	14.3 (7.8)	19.1 (8.5)	<0.001
GCS[a]	102	14.1 (2.1)	13.6 (2.6)	14.1 (2.1)	13.2 (3.1)	<0.001
Comorbidity of[b]						
CHF	0	149 (11.5)	146 (10.7)	577 (20.0)	408 (42.1)	<0.001
AFIB	0	309 (23.8)	222 (16.3)	626 (21.8)	344 (35.5)	<0.001
CRD	0	67 (5.2)	59 (4.3)	179 (6.2)	383 (39.6)	<0.001
Liver disease	0	42 (3.2)	144 (10.6)	129 (4.5)	96 (9.9)	<0.001
COPD	0	181 (13.9)	153 (11.2)	487 (16.9)	167 (17.3)	<0.001
CAD	0	196 (15.1)	182 (13.4)	386 (13.4)	232 (24.0)	<0.001
Stroke	0	29 (2.2)	206 (15.1)	118 (4.1)	30 (3.1)	<0.001
Non-metastatic solid cancer[b]	0	688 (53.0)	565 (41.5)	731 (25.4)	290 (30.0)	<0.001
Metastatic cancer[b]	0	525 (40.4)	646 (47.5)	1,340 (46.6)	360 (37.2)	<0.001
Hematologic malignancy[b]	0	94 (7.2)	153 (11.2)	822 (28.6)	324 (33.5)	<0.001
Vasopressor use[b]	0	403 (30.8)	320 (23.2)	469 (16.8)	544 (52.6)	<0.001
Mechanical ventilation use[b]	0	653 (50.0)	740 (53.7)	745 (26.8)	512 (49.5)	<0.001
RRT use[b]		5 (0.4)	4 (0.3)	11 (0.4)	152 (15.7)	<0.001

Table 1 (*continued*)

Table 1 (*continued*)

	Missing	Cluster 1	Cluster 2	Cluster 3	Cluster 4	P value
Heart rate[a] (bpm)	87	86.6 (14.2)	84.2 (15.5)	92.1 (17.1)	92.4 (18.2)	<0.001
Mean blood pressure[a] (mmHg)	89	78.1 (9.5)	81.3 (10.6)	77.4 (10.8)	72.1 (11.1)	<0.001
Temperature[a] (℃)	137	36.9 (0.6)	36.8 (0.6)	36.8 (0.6)	36.5 (0.7)	<0.001
Oxygen saturation fraction[a] (%)	98	97.5 (1.7)	97.6 (1.7)	96.6 (2.0)	95.7 (4.5)	<0.001
CVP[a] (mmHg)	4,927	9.7 (5.2)	9.8 (5.0)	10.8 (6.1)	16.8 (8.2)	<0.001
Albumin[a] (g/dL)	3,834	2.8 (0.6)	3.1 (0.7)	3.0 (0.6)	2.7 (0.6)	<0.001
Bilirubin[a] (mg/dL)	3,277	2.2 (2.9)	1.9 (2.9)	1.7 (3.3)	3.5 (6.3)	<0.001
INR[a]	877	1.3 (0.3)	1.3 (0.4)	1.4 (0.6)	2.0 (1.7)	<0.001
Creatinine[a] (mg/dL)	56	1.0 (0.7)	0.9 (0.5)	0.9 (0.5)	2.9 (2.4)	<0.001
BUN[a] (mg/dL)	59	18.4 (10.0)	18.9 (10.0)	21.3 (12.4)	54.0 (28.6)	<0.001
Hematocrit[a] (%)	59	32.2 (4.4)	32.8 (5.3)	30.3 (5.2)	30.4 (5.3)	<0.001
WBC[a] (1,000/uL)	78	12.4 (6.1)	11.9 (7.9)	12.0 (16.4)	19.2 (41.0)	<0.001
Platelet[a] (1,000/uL)	76	232.5 (110.5)	242.5 (129.8)	235.3 (161.4)	195.6 (131.6)	<0.001
Lactate[a] (mmol/L)	2,460	2.2 (1.4)	2.3 (1.5)	1.9 (1.0)	3.8 (3.3)	<0.001
Glucose[a] (mg/dL)	43	143.4 (32.5)	143.8 (43.5)	134.1 (47.4)	150.5 (73.1)	<0.001
Potassium[a] (mmol/L)	48	4.2 (0.4)	4.1 (0.4)	4.0 (0.5)	4.7 (0.8)	<0.001
Sodium[a] (mmol/L)	56	137.8 (3.0)	137.7 (4.0)	137.4 (4.7)	136.9 (6.0)	<0.001
pH[a]	3,813	7.4 (0.1)	7.4 (0.1)	7.4 (0.1)	7.3 (0.1)	<0.001
PaO_2[a] (mmHg)	3,813	171.4 (66.8)	211.9 (89.7)	127.2 (59.8)	130.4 (62.6)	<0.001
PCO_2[a] (mmHg)	3,813	42.1 (6.9)	40.4 (7.2)	43.3 (12.4)	40.8 (11.9)	<0.001
PaO_2/FiO_2[a]	3,813	291.4 (120.9)	344.5 (148.3)	217.5 (110.6)	203.3 (110.4)	<0.001
Length of hospital stay[a] (d)	0	3.8 (6.1)	3.6 (4.6)	3.6 (5.1)	5.1 (7.2)	<0.001
Hospital death[a]	0	56 (4.3)	162 (11.7)	550 (19.7)	481 (46.5)	<0.001

[a], continued value summarized as mean (SD); [b], categorical value summarized as number (percentage). CCU, cardiac care unit; CSRU, Cardiac surgery recovery unit; MICU medical intensive care unit; SICU, surgery intensive care unit; TSICU, trauma/surgical intensive care unit; SOFA, sequential organ failure assessment; SAPS Simplified Acute Physiology Score; GCS, Glasgow Coma Scale; CHF, chronic heart failure; AFIB atrial fibrillation; CRD chronic renal disease; COPD, chronic obstructive pulmonary disease; CAD, coronary artery disease; RRT, renal replacement treatment; CVP, central venal pressure; INR, international normalized ratio; BUN, blood urea nitrogen; WBC, white blood cells; PaO_2, artery partial pressure of oxygen; PCO_2, partial pressure of carbon dioxide; FiO_2, fraction of inspired oxygen.

factors (23,24). In the SAPS II system, metastatic cancer was assigned an increasing score of 9 whereas hematologic malignancy was scored as 10 (14).

Generally, the primary diagnosis is always the main reason for ICU admission. As shown in *Figure 1*, most non-metastatic solid cancer held the first position of diagnosis. And in *Figure 2*, non-metastatic solid cancer more frequently occurs in cluster 1 and 2, which were featured as surgical patients. All the above imply that these patients were admitted to ICU for postoperative monitoring. On the contrary, metastatic cancer and hematologic malignancy are less likely to be the primary diagnosis, which implies that its complications, rather than the cancer itself, necessitated ICU admission for these patients.

We found that different clinical subtypes could be simply identified by the admission information. Hospital mortality varied among these subtypes, and this was consistent with the variation of prognoses reported by previous studies

Table 2 The standard mortality ratio and Hosmer–Lemeshow C statistic of different groups by different severity-scoring systems

	SMR (95% CIs)	H–L C statistic	P value
Cluster 1+2 (MIMIC) to SAPS II	0.37 (0.33–0.43)	409.22	<0.001
Cluster 3+4 (MIMIC) to SAPS II	0.76 (0.72–0.81)	315.15	<0.001
Whole population (MIMIC) to SAPS II	0.65 (0.61–0.68)	628.07	<0.001
Surgical group (eICU) to SAPS II	0.32 (0.25–0.39	209.53	<0.001
Non-surgical group (eICU) to SAPS II	0.67 (0.63–0.71)	679.19	<0.001
Whole population (eICU) to SAPS II	0.63 (0.59–0.67)	827.64	<0.001
Surgical group (eICU) to APACHE IV	0.76 (0.60–0.95)	31.53	<0.001
Non-surgical group (eICU) to APACHE IV	1.07 (1.01–1.14)	99.19	<0.001
Whole population (eICU) to APACHE IV	1.05 (0.99–1.11)	87.84	<0.001

SMR, standard mortality ratio; H-L Hosmer–Lemeshow; MIMIC, medical information mart for intensive care; SPAS, Simplified Acute Physiology Score; APACHE, acute physiology and chronic health evaluation.

(6,25). The reported hospital mortality rates could be as low as 10%, often in postoperative monitoring cohorts (4,8), and as high as 50% or more, often in non-surgical diagnoses (3,9). This disparity cannot be explained only by illness severity but may indicate a different association of mortality other than that mentioned in the traditional severity scoring systems. One reason for this disparity may be that the acute postoperative stress status (tachycardia, intubation, and so on) causes a deterioration of the score evaluated in the first 24 hours in the ICU. However, more research is needed to validate this rationale.

The SAPS II scoring system didn't perform well in calibration even though it considered the influence of admission type, partially due to its older age (introduced in 1993). The APACHE IV system takes the admission type and diagnosis into account when predicting mortality according to the APACHE score (26). The system performed well for cancer patients at low risk of mortality in the eICU database of the present study. However, as the number of ICU cancer patients at high risk of mortality in the present study (especially the surgical group) was too low to generate credible conclusions, more evidence is needed. Previous studies have found an apriority of APACHE IV compared to SAPS II in many ICU conditions (27,28), but the evidence in cancer patients was insufficient. The application of the APACHE IV system is limited by its complexity. However, as data science develops, the complexity of this system could be overcome by electronic recording techniques and massive computing power.

The present study supports the theory that cancer patients in the ICU need to be considered separately according to the admission type and clinical service provider. A similar opinion was proposed by Weissman and his colleagues based on their experience and intuition (29). They found significant differences between elective and emergency postoperative patients in the ICU, which corresponds to the first and second clusters in our study. However, our study used a method of machine learning to deal with the high-dimensional clinical data, based on the theory that data intraclusters were more consistent than interclusters. Furthermore, The K-mean algorithm tells us the distance among different clusters. When K=2 we have surgical and medical groups; until K=4 we get elective and emergent/urgent subgroups. It implies that the differences between surgical and medical groups are more obvious than those between elective and emergent/urgent subgroups. Separately analysing the surgical and medical patients (or elective and emergent/urgent surgical) with cancer in ICU can decrease heterogeneity. Some studies addressed this concern and specified patient-selection criteria (7,30,31), but others lacked this additional measure and only treated it as a variable for analysis (1,2,32). This is because there are some other concerns (e.g., sample size, study purpose) to take into account besides heterogeneity for observational research as well as systematic review or meta-analysis.

Puxty and colleagues conducted excellent work on a comprehensive meta-analysis for ICU patients with solid cancers (25). They found that cancer patients admitted as medical patients had increased risk (2- to 4-fold higher) of ICU mortality, and increased risk of in-hospital mortality (6- to 8-fold higher) compared with cancer patients who were surgical admissions; however, they only specified

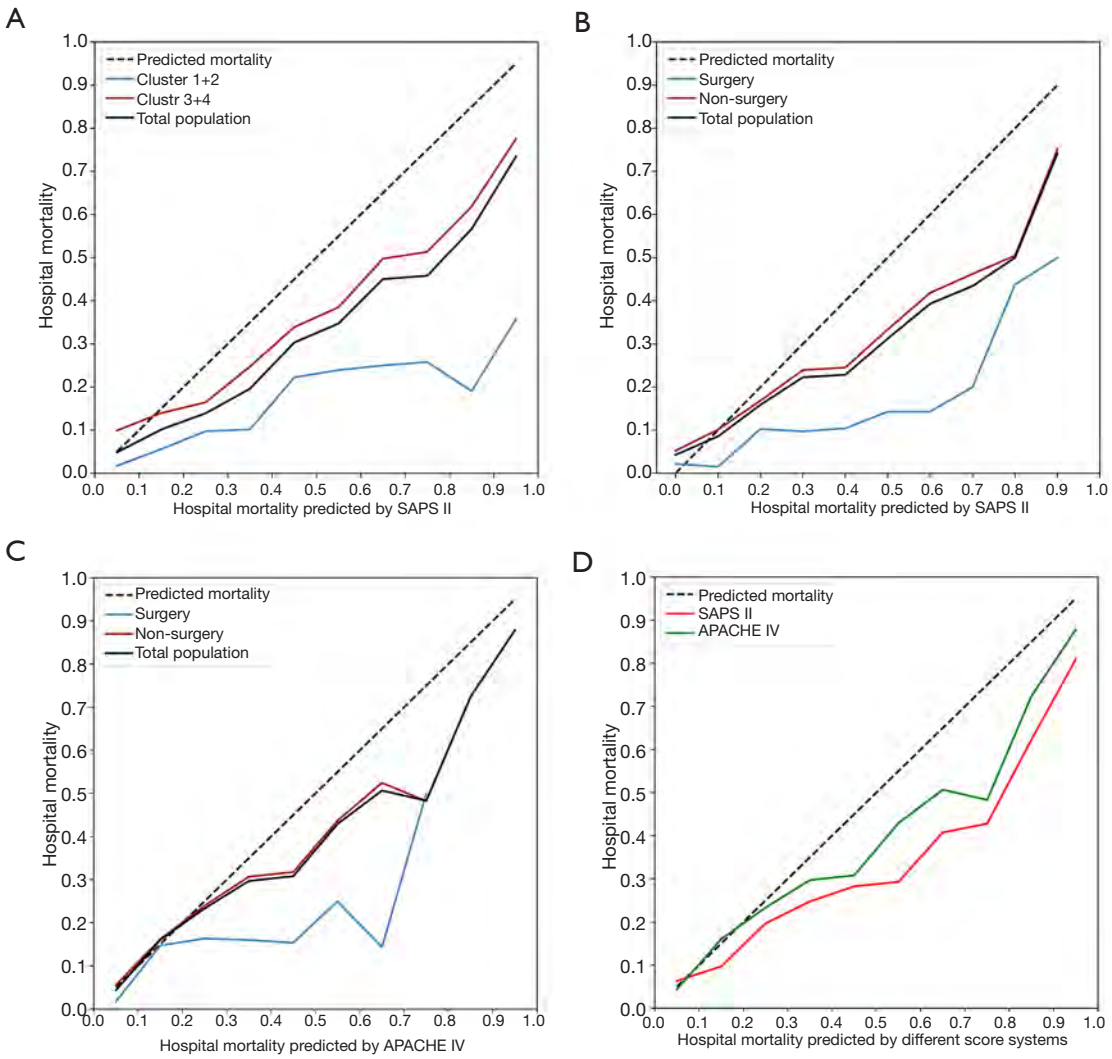

Figure 3 The graphs show the calibration of SAPS II and APACHE IV in the MIMIC and eICU databases: (A) SAPS II in the MIMIC database; (B) SAPS II in the eICU database; (C) APACHE IV in the eICU database; (D) the comparison of the total population for the two score systems in eICU database. MIMIC, Medical Information Mart for Intensive Care; SPAS, Simplified Acute Physiology Score; APACHE, Acute Physiology and Chronic Health Evaluation.

group stratification by different cancer sites. Furthermore, we noticed an outstanding variation among the enrolled studies for both hospital mortality (I^2=98.8%, P<0.001) and ICU mortality (I^2=99.1%, P<0.001). Therefore, the admission type should have been considered for stratified analysis in Puxty's research, which may have also improved the heterogeneity of the enrolled studies.

Our study provided ICD-level data on the mortality rate of cancer patients in the ICU from multiple resources, and emphasized the importance of differentiating between admission types and clinical service providers. We discussed the calibration of different severity-scoring systems for cancer patients in the ICU across different clinical subgroups. However, this study has some limitations. First, selection bias might exist, especially in the eICU database, because there were fewer cancer patients in the ICU than reported on average by most studies (~15%). Second, unlike the eICU database, it was difficult to compare the two different severity-scoring systems in the MIMIC database, because the APACHE IV algorithm is only available online (33). Moreover, there were limits deriving from the retrospective nature of the analysis and relatively small

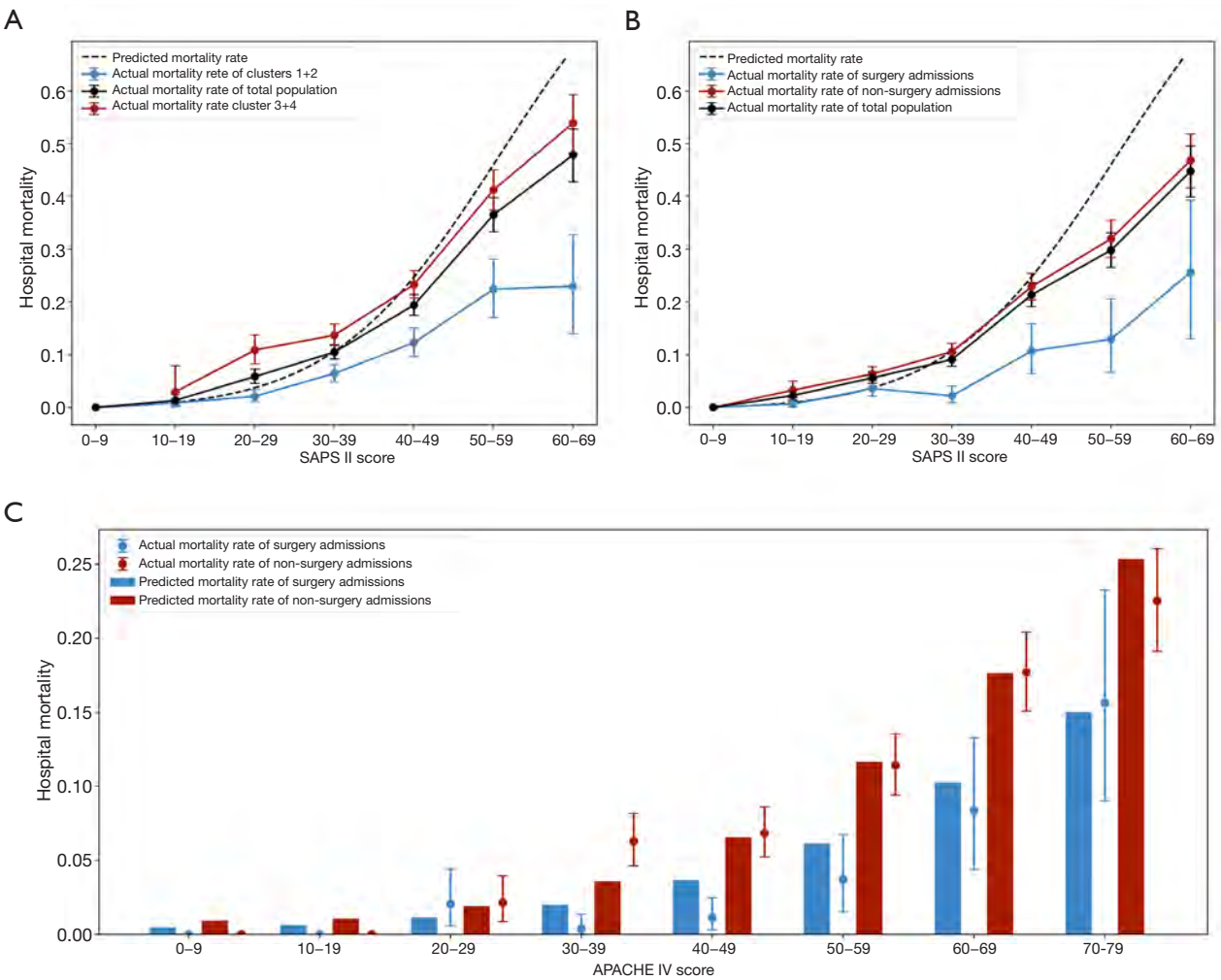

Figure 4 The comparison of mortality in different groups stratified by different intervals of severity score. (A) The population from the MIMIC database; (B,C) the population from the eICU database. MIMIC, Medical Information Mart for Intensive Care; SPAS, Simplified Acute Physiology Score; APACHE, Acute Physiology and Chronic Health Evaluation.

sample size for machine learning. Therefore, prospective studies with larger populations are needed in the future.

Conclusions

Similar mortality rates for a spectrum of ICD-level cancer diagnoses were observed within two large ICU databases. Different clinical subtypes of ICU patients with cancer can be well identified by admission type and clinical service provider, and caution should be exercised when considering these patients as a whole population, both in clinical practice and research. The system of APACHE IV performs better than SAPS II for cancer patients at low risk of mortality in the ICU, whereas the same conclusion for those at high risk of mortality needs to be validated in further studies.

Data sharing and data accessibility

The datasets generated and/or analysed during the current study are available in the MIMIC-III v1.4 (https://mimic.physionet.org) and eICU (https://eicu-crd.mit.edu/).

Acknowledgments

We appreciate the researchers at the MIT Laboratory for Computational Physiology for publicly sharing of the clinical databases including Medical Information Mart for Intensive Care-III and eICU database.

Funding: The study was funded by the National Natural Science Foundation of China (No. 81770619).

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at http://dx.doi.org/10.21037/atm-20-4634

Peer Review File: Available at http://dx.doi.org/10.21037/atm-20-4634

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-4634). All the authors report grants from National Natural Science Foundation of China, during the conduct of the study. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Soares M, Caruso P, Silva E, et al. Characteristics and outcomes of patients with cancer requiring admission to intensive care units: a prospective multicenter study. Crit Care Med 2010;38:9-15.
2. Taccone FS, Artigas AA, Sprung CL, et al. Characteristics and outcomes of cancer patients in European ICUs. Crit Care 2009;13:R15.
3. Caruso P, Ferreira AC, Laurienzo CE, et al. Short- and long-term survival of patients with metastatic solid cancer admitted to the intensive care unit: prognostic factors. Eur J Cancer Care (Engl) 2010;19:260-6.
4. Leath CA 3rd, Kendrick JE 4th, Numnum TM, et al. Outcomes of gynecologic oncology patients admitted to the intensive care unit following surgery: A university teaching hospital experience. Int J Gynecol Cancer 2006;16:1766-9.
5. Darmon M, Bourmaud A, Georges Q, et al. Changes in critically ill cancer patients' short-term outcome over the last decades: results of systematic review with meta-analysis on individual data. Intensive Care Med 2019;45:977-87.
6. Shimabukuro-Vornhagen A, Böll B, Kochanek M, et al. Critical care of patients with cancer. CA Cancer J Clin 2016;66:496-517.
7. Puxty K, McLoone P, Quasim T, et al. Characteristics and Outcomes of Surgical Patients With Solid Cancers Admitted to the Intensive Care Unit. JAMA Surg 2018;153:834-40.
8. Park DP, Welch CA, Harrison DA, et al. Outcomes following oesophagectomy in patients with oesophageal cancer: a secondary analysis of the ICNARC Case Mix Programme Database. Crit Care 2009;13 Suppl 2:S1.
9. Reichner CA, Thompson JA, O'Brien S, et al. Outcome and code status of lung cancer patients admitted to the medical ICU. Chest 2006;130:719-23.
10. Ishida T, Inoue T, Sugio K, et al. Early squamous lung cancer and longer survival rates. Respiration 1993;60:359-65.
11. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016;3:160035.
12. Pollard TJ, Johnson AEW, Raffa JD, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data 2018;5:180178.
13. Feng M, McSparron JI, Kien DT, et al. Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database. Intensive Care Med 2018;44:884-92.
14. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270:2957-63.
15. van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009;47:626-33.
16. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med 1996;22:707-10.
17. mimic-code. Accessed Feb 4th 2020. Available online: https://github.com/MIT-LCP/mimic-code/tree/master/concepts/severityscore

18. Yadav J, Sharma M. A Review of K-mean Algorithm. Int J Latest Trends Eng Technol 2013;4:2972-6.
19. Kodinariya TM, Makwana PR. Review on determining number of Cluster in K-Means Clustering. Int J Adv Res Comp Sci Manag Stud 2013;1:90-5.
20. Ivosev G, Burton L, Bonner R. Dimensionality reduction and visualization in principal component analysis. Anal Chem 2008;80:4933-44.
21. Hosmer DW, Hosmer T, Le Cessie S, et al. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997;16:965-80.
22. Vandenbroucke JP. A shortcut method for calculating the 95 per cent confidence interval of the standardized mortality ratio. Am J Epidemiol 1982;115:303-4.
23. Mendoza V, Lee A, Marik PE. The hospital-survival and prognostic factors of patients with solid tumors admitted to an ICU. Am J Hosp Palliat Care 2008;25:240-3.
24. Oeyen SG, Benoit DD, Annemans L, et al. Long-term outcomes and quality of life in critically ill patients with hematological or solid malignancies: a single center study. Intensive Care Med 2013;39:889-98.
25. Puxty K, McLoone P, Quasim T, et al. Survival in solid cancer patients following intensive care unit admission. Intensive Care Med 2014;40:1409-28.
26. Zimmerman JE, Kramer AA, McNair DS, et al. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006;34:1297-310.
27. Lee H, Yoon S, Oh SY, et al. Comparison of APACHE IV with APACHE II, SAPS 3, MELD, MELD-Na, and CTP scores in predicting mortality after liver transplantation. Sci Rep 2017;7:10884.
28. Norouzi K, Mashmool Z, Dalvandi A, et al. Comparison of two tools APACHE IV and SAPS II in predicting mortality rate in patients hospitalized in intensive care unit. Koomesh 2015;16:347-55.
29. Weissman C, Klein N. The importance of differentiating between elective and emergency postoperative critical care patients. J Crit Care 2008;23:308-16.
30. Aygencel G, Turkoglu M, Turkoz Sucak G, et al. Prognostic factors in critically ill cancer patients admitted to the intensive care unit. J Crit Care 2014;29:618-26.
31. Ostermann M, Ferrando-Vivas P, Gore C, et al. Characteristics and Outcome of Cancer Patients Admitted to the ICU in England, Wales, and Northern Ireland and National Trends Between 1997 and 2013. Crit Care Med 2017;45:1668-76.
32. Bos MMEM, Bakhshi-Raiez F, Dekker JWT, et al. Outcomes of intensive care unit admissions after elective cancer surgery. Eur J Surg Oncol 2013;39:584-92.
33. Apache IV Score. Accessed Feb 16th 2020. Available online: https://intensivecarenetwork.com/Calculators/Files/Apache4.html

Cite this article as: Gao S, Wang Y, Yang L, Wang Z, Huang W. Characteristics and clinical subtypes of cancer patients in the intensive care unit: a retrospective observational study for two large databases. Ann Transl Med 2021;9(1):13. doi: 10.21037/atm-20-4634

Supplementary

Table S1 The top-10 most common and most fatal types of cancer in MIMIC database

Cancer type	Count	Death
Top-10 most common cancer type		
Secondary malignant neoplasm of bone and bone marrow	799	201
Malignant neoplasm of liver, secondary	758	237
Secondary malignant neoplasm of lung	723	184
Secondary malignant neoplasm of brain and spinal cord	703	140
Secondary malignant neoplasm of other specified sites	423	80
Malignant neoplasm of other parts of bronchus or lung	370	99
Malignant neoplasm of liver, primary	329	68
Other malignant lymphomas, unspecified site, extranodal and solid organ sites	315	66
Malignant neoplasm of upper lobe, bronchus or lung	309	51
Secondary malignant neoplasm of retroperitoneum and peritoneum	299	79
Top-10 most fatal cancer type		
Mycosis fungoides, unspecified site, extranodal and solid organ sites	22	11
Acute myeloid leukemia, in relapse	13	6
Secondary and unspecified malignant neoplasm of lymph nodes of multiple sites	34	15
Acute myeloid leukemia, without mention of having achieved remission	189	78
Secondary malignant neoplasm of pleura	207	78
Malignant neoplasm of pancreas, part unspecified	32	12
Reticulosarcoma, unspecified site, extranodal and solid organ sites	17	6
Acute lymphoid leukemia, without mention of having achieved remission	38	13
Malignant neoplasm of middle lobe, bronchus or lung	30	10
Other malignant neoplasm without specification of site	151	50

http://dx.doi.org/10.21037/atm-20-4634

Table S2 The top-10 most common and most fatal types of cancer in eICU database

Top-10 most common cancer type	Count	Death
oncology|chest tumors|primary lung cancer	1377	294
neurologic|CNS mass lesions|brain tumor	1065	79
oncology|chest tumors|metastatic lung CA	477	132
oncology|GI tumors|colon CA	411	61
oncology|GU tumors|prostate CA	326	43
oncology|GI tumors|liver CA|CA metastatic to liver	250	68
oncology|chest tumors|breast CA|female	234	40
oncology|GI tumors|esophageal CA	226	33
oncology|hematologic malignancy|lymphoproliferative disease|non-Hodgkin's lymphoma	219	50
oncology|GU tumors|renal cell CA|right kidney	202	33
Top-10 most fatal cancer type	Count	Death
oncology|unknown primary|metastatic adenocarcinoma|unknown primary	10	5
hematology|oncology and leukemia|leukemia|acute myelogenous	137	52
oncology|GI tumors|colon CA|cecum	20	7
oncology|hematologic malignancy|leukemia|acute lymphocytic	36	12
pulmonary|pleural disorders|pleural effusion|malignant	113	36
oncology|chest tumors|mediastinal tumor|lymphoma	45	13
oncology|skin, muscle and skeletal tumors|bone tumors|bony metastasis	142	41
oncology|chest tumors|metastatic lung CA	477	132
oncology|GI tumors|liver CA|CA metastatic to liver	250	68
hematology|oncology and leukemia|lymphoproliferative disease|Hodgkin's disease	43	11

CNS central nervous system, GI gastrointestinal, GU Genitourinary, CA cancer

http://dx.doi.org/10.21037/atm-20-4634

Table S3 Manually classified cancer list of MIMIC database according to the belonging system or location

ICD9 code	Long title	Manually classification
1504	Malignant neoplasm of middle third of esophagus	Digestive
1505	Malignant neoplasm of lower third of esophagus	Digestive
1508	Malignant neoplasm of other specified part of esophagus	Digestive
1509	Malignant neoplasm of esophagus, unspecified site	Digestive
1510	Malignant neoplasm of cardia	Digestive
1512	Malignant neoplasm of pyloric antrum	Digestive
1514	Malignant neoplasm of body of stomach	Digestive
1518	Malignant neoplasm of other specified sites of stomach	Digestive
1519	Malignant neoplasm of stomach, unspecified site	Digestive
1520	Malignant neoplasm of duodenum	Digestive
1531	Malignant neoplasm of transverse colon	Digestive
1532	Malignant neoplasm of descending colon	Digestive
1533	Malignant neoplasm of sigmoid colon	Digestive
1534	Malignant neoplasm of cecum	Digestive
1536	Malignant neoplasm of ascending colon	Digestive
1538	Malignant neoplasm of other specified sites of large intestine	Digestive
1539	Malignant neoplasm of colon, unspecified site	Digestive
1540	Malignant neoplasm of rectosigmoid junction	Digestive
1541	Malignant neoplasm of rectum	Digestive
1548	Malignant neoplasm of other sites of rectum, rectosigmoid junction, and anus	Digestive
1550	Malignant neoplasm of liver, primary	Digestive
1551	Malignant neoplasm of intrahepatic bile ducts	Digestive
1552	Malignant neoplasm of liver, not specified as primary or secondary	Digestive
1560	Malignant neoplasm of gallbladder	Digestive
1561	Malignant neoplasm of extrahepatic bile ducts	Digestive
1562	Malignant neoplasm of ampulla of vater	Digestive
1570	Malignant neoplasm of head of pancreas	Digestive
1572	Malignant neoplasm of tail of pancreas	Digestive
1578	Malignant neoplasm of other specified sites of pancreas	Digestive
1579	Malignant neoplasm of pancreas, part unspecified	Digestive
1580	Malignant neoplasm of retroperitoneum	Digestive
1588	Malignant neoplasm of specified parts of peritoneum	Digestive
1619	Malignant neoplasm of larynx, unspecified	Respiratory
1622	Malignant neoplasm of main bronchus	Respiratory
1623	Malignant neoplasm of upper lobe, bronchus or lung	Respiratory
1624	Malignant neoplasm of middle lobe, bronchus or lung	Respiratory
1625	Malignant neoplasm of lower lobe, bronchus or lung	Respiratory
1628	Malignant neoplasm of other parts of bronchus or lung	Respiratory
1629	Malignant neoplasm of bronchus and lung, unspecified	Respiratory
1725	Malignant melanoma of skin of trunk, except scrotum	Skin
1729	Melanoma of skin, site unspecified	Skin
1733	Oth and unspec malignant neoplasm of skin	Skin
1748	Malignant neoplasm of other specified sites of female breast	Breast and thyroid
1749	Malignant neoplasm of breast (female), unspecified	Breast and thyroid
1820	Malignant neoplasm of corpus uteri, except isthmus	Genitourinary
1830	Malignant neoplasm of ovary	Genitourinary
185	Malignant neoplasm of prostate	Genitourinary
1888	Malignant neoplasm of other specified sites of bladder	Genitourinary
1889	Malignant neoplasm of bladder, part unspecified	Genitourinary
1890	Malignant neoplasm of kidney, except pelvis	Genitourinary
1911	Malignant neoplasm of frontal lobe	Brain
1912	Malignant neoplasm of temporal lobe	Brain
1913	Malignant neoplasm of parietal lobe	Brain
1918	Malignant neoplasm of other parts of brain	Brain
1919	Malignant neoplasm of brain, unspecified	Brain
193	Malignant neoplasm of thyroid gland	Breast and thyroid
1960	Secondary and unspecified malignant neoplasm of lymph nodes of head, face, and neck	Secondary
1961	Secondary and unspecified malignant neoplasm of intrathoracic lymph nodes	Secondary
1962	Secondary and unspecified malignant neoplasm of intra-abdominal lymph nodes	Secondary
1963	Secondary and unspecified malignant neoplasm of lymph nodes of axilla and upper limb	Secondary
1965	Secondary and unspecified malignant neoplasm of lymph nodes of inguinal region and lower limb	Secondary
1966	Secondary and unspecified malignant neoplasm of intrapelvic lymph nodes	Secondary
1968	Secondary and unspecified malignant neoplasm of lymph nodes of multiple sites	Secondary
1969	Secondary and unspecified malignant neoplasm of lymph nodes, site unspecified	Secondary
1970	Secondary malignant neoplasm of lung	Secondary
1971	Secondary malignant neoplasm of mediastinum	Secondary
1972	Secondary malignant neoplasm of pleura	Secondary
1973	Secondary malignant neoplasm of other respiratory organs	Secondary
1974	Secondary malignant neoplasm of small intestine including duodenum	Secondary
1975	Secondary malignant neoplasm of large intestine and rectum	Secondary
1976	Secondary malignant neoplasm of retroperitoneum and peritoneum	Secondary
1977	Malignant neoplasm of liver, secondary	Secondary
1978	Secondary malignant neoplasm of other digestive organs and spleen	Secondary
1980	Secondary malignant neoplasm of kidney	Secondary
1981	Secondary malignant neoplasm of other urinary organs	Secondary
1982	Secondary malignant neoplasm of skin	Secondary
1983	Secondary malignant neoplasm of brain and spinal cord	Secondary
1984	Secondary malignant neoplasm of other parts of nervous system	Secondary
1985	Secondary malignant neoplasm of bone and bone marrow	Secondary
1986	Secondary malignant neoplasm of ovary	Secondary
1987	Secondary malignant neoplasm of adrenal gland	Secondary
19882	Secondary malignant neoplasm of genital organs	Secondary
19889	Secondary malignant neoplasm of other specified sites	Secondary
1991	Other malignant neoplasm without specification of site	Secondary
20000	Reticulosarcoma, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20020	Burkitt's tumor or lymphoma, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20080	Other named variants of lymphosarcoma and reticulosarcoma, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20190	Hodgkin's disease, unspecified type, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20200	Nodular lymphoma, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20210	Mycosis fungoides, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20240	Leukemic reticuloendotheliosis, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20280	Other malignant lymphomas, unspecified site, extranodal and solid organ sites	Hematologic malignancy
20281	Other malignant lymphomas, lymph nodes of head, face, and neck	Hematologic malignancy
20283	Other malignant lymphomas, intra-abdominal lymph nodes	Hematologic malignancy
20288	Other malignant lymphomas, lymph nodes of multiple sites	Hematologic malignancy
20300	Multiple myeloma, without mention of having achieved remission	Hematologic malignancy
20400	Acute lymphoid leukemia, without mention of having achieved remission	Hematologic malignancy
20410	Chronic lymphoid leukemia, without mention of having achieved remission	Hematologic malignancy
20500	Acute myeloid leukemia, without mention of having achieved remission	Hematologic malignancy
20501	Acute myeloid leukemia, in remission	Hematologic malignancy
20502	Acute myeloid leukemia, in relapse	Hematologic malignancy
20510	Chronic myeloid leukemia, without mention of having achieved remission	Hematologic malignancy

http://dx.doi.org/10.21037/atm-20-4634

Table S4 Manually classified cancer list of eICU database according to the belonging system or location

ICD9 code	Diagnoses	Manually classification
145.9	oncology|head and neck tumors|mouth and jaw tumor|oral cavity squamous cell CA	Head and neck
148.9	oncology|head and neck tumors|neck tumor|oro/hypopharyngeal CA	Head and neck
151.9	oncology|GI tumors|gastric CA	Digestive
152.9	oncology|GI tumors|intestinal CA	Digestive
153	oncology|GI tumors|colon CA|right hemicolon	Digestive
153.3	oncology|GI tumors|colon CA|sigmoid	Digestive
153.4	oncology|GI tumors|colon CA|cecum	Digestive
153.9	oncology|GI tumors|colon CA	Digestive
154.3	oncology|GI tumors|anal CA	Digestive
155	oncology|GI tumors|liver CA|hepatocellular CA	Digestive
156.9	oncology|GI tumors|cholangiocarcinoma	Digestive
157.9	oncology|GI tumors|pancreatic tumor|pancreatic CA	Digestive
161.9	oncology|head and neck tumors|neck tumor|laryngeal CA	Head_neck
162.9	oncology|chest tumors|primary lung cancer	Lung
170.9	oncology|skin, muscle and skeletal tumors|bone tumors	Skin, muscle and skeletal
171.9	oncology|skin, muscle and skeletal tumors|soft tissue sarcoma	Skin, muscle and skeletal
172.9	oncology|skin, muscle and skeletal tumors|melanoma	Skin, muscle and skeletal
173.9	oncology|skin, muscle and skeletal tumors|basal cell CA	Skin, muscle and skeletal
174.9	oncology|chest tumors|breast CA|female	Breast and thyroid
179	oncology|GU tumors|uterine CA	Genitourinary
180.9	oncology|GU tumors|cervical CA	Genitourinary
183	oncology|GU tumors|ovarian CA	Genitourinary
185	oncology|GU tumors|prostate CA	Genitourinary
186.9	oncology|GU tumors|testicular CA	Genitourinary
188.9	oncology|GU tumors|bladder CA	Genitourinary
189	oncology|GU tumors|renal cell CA|right kidney	Genitourinary
191.9	neurologic|CNS mass lesions|brain tumor	Brain
192.2	neurologic|CNS mass lesions|spinal tumor	Brain
192.9	oncology|GI tumors|esophageal CA	Brain
192.9	oncology|CNS tumors|spinal cord tumors	Brain
193	oncology|head and neck tumors|neck tumor|thyroid tumor|papillary CA	Breast and thyroid
197	oncology|chest tumors|metastatic lung CA	Secondary
197.2	pulmonary|pleural disorders|pleural effusion|malignant	Secondary
197.6	oncology|GI tumors|abdominal carcinomatosis	Secondary
197.7	oncology|GI tumors|liver CA|CA metastatic to liver	Secondary
198.3	neurologic|CNS mass lesions|brain tumor|metastatic brain tumor	Secondary
198.5	oncology|skin, muscle and skeletal tumors|bone tumors|bony metastasis	Secondary
199	oncology|unknown primary|metastatic adenocarcinoma	Secondary
199.1	oncology|unknown primary|metastatic adenocarcinoma|unknown primary	Secondary
201.9	hematology|oncology and leukemia|lymphoproliferative disease|Hodgkin's disease	Hematologic malignancy
202.8	oncology|hematologic malignancy|lymphoproliferative disease|non-Hodgkin's lymphoma	Hematologic malignancy
202.82	oncology|chest tumors|mediastinal tumor|lymphoma	Hematologic malignancy
203	oncology|hematologic malignancy|multiple myeloma	Hematologic malignancy
204	oncology|hematologic malignancy|leukemia|acute lymphocytic	Hematologic malignancy
204.1	oncology|hematologic malignancy|leukemia|chronic lymphocytic	Hematologic malignancy
205	hematology|oncology and leukemia|leukemia|acute myelogenous	Hematologic malignancy
205.1	oncology|hematologic malignancy|leukemia|chronic myelogenous	Hematologic malignancy

CNS, central nervous system; GI, gastrointestinal; GU, Genitourinary, CA, cancer.

http://dx.doi.org/10.21037/atm-20-4634

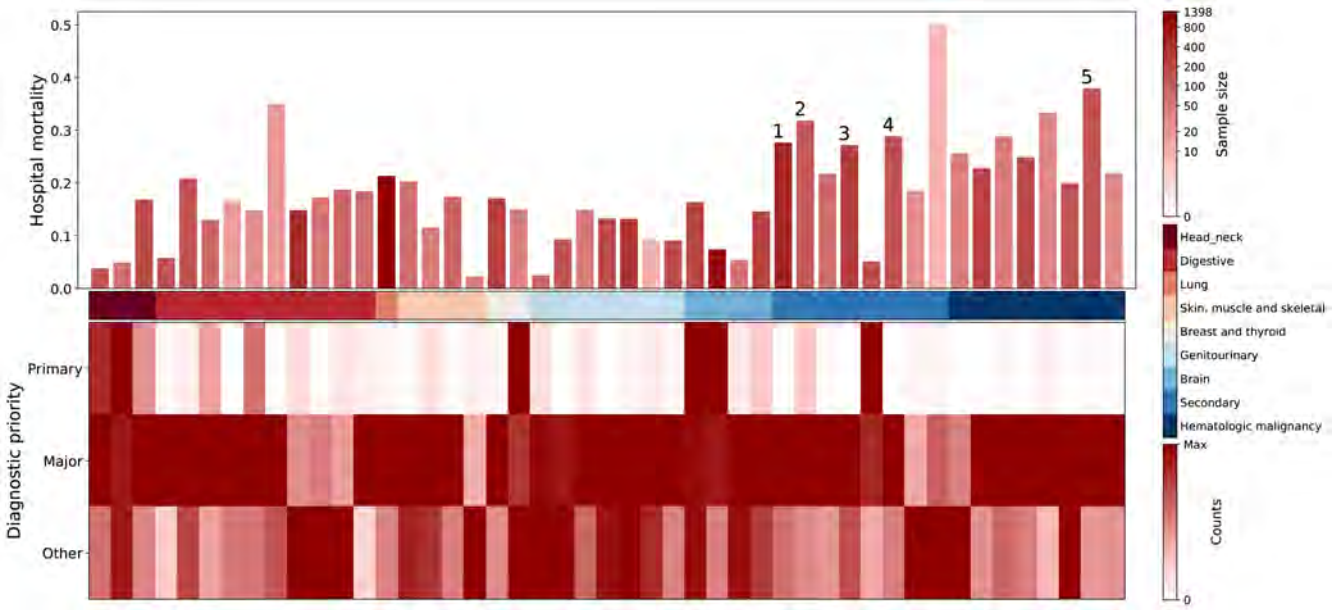

Figure S1 The mortality and distribution of positions of ICD-level diagnoses of EICU database. In the upper part, the color of bars referred to the number of patients with specific cancer. In the middle part, the color blocks referred to the manually classified according to the location or organ system. In the lower part was the heatmap. The color in the primary/major/other diagnosis referred to its frequency. The note label of figure above were: 1, metastatic lung cancer; 2, malignant pleural disorders; 3, cancer metastatic to liver; 4, bony metastasis; 5, acute myeloid leukemia.

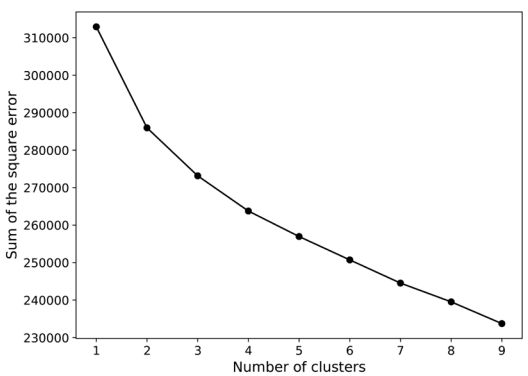

Figure S2 The elbow plot of K-means clustering of the ICU cancer patient in MIMIC database. The sum of the square error was calculated by summing within-cluster Euclidian distance of every observation to its nearest cluster center.

http://dx.doi.org/10.21037/atm-20-4634

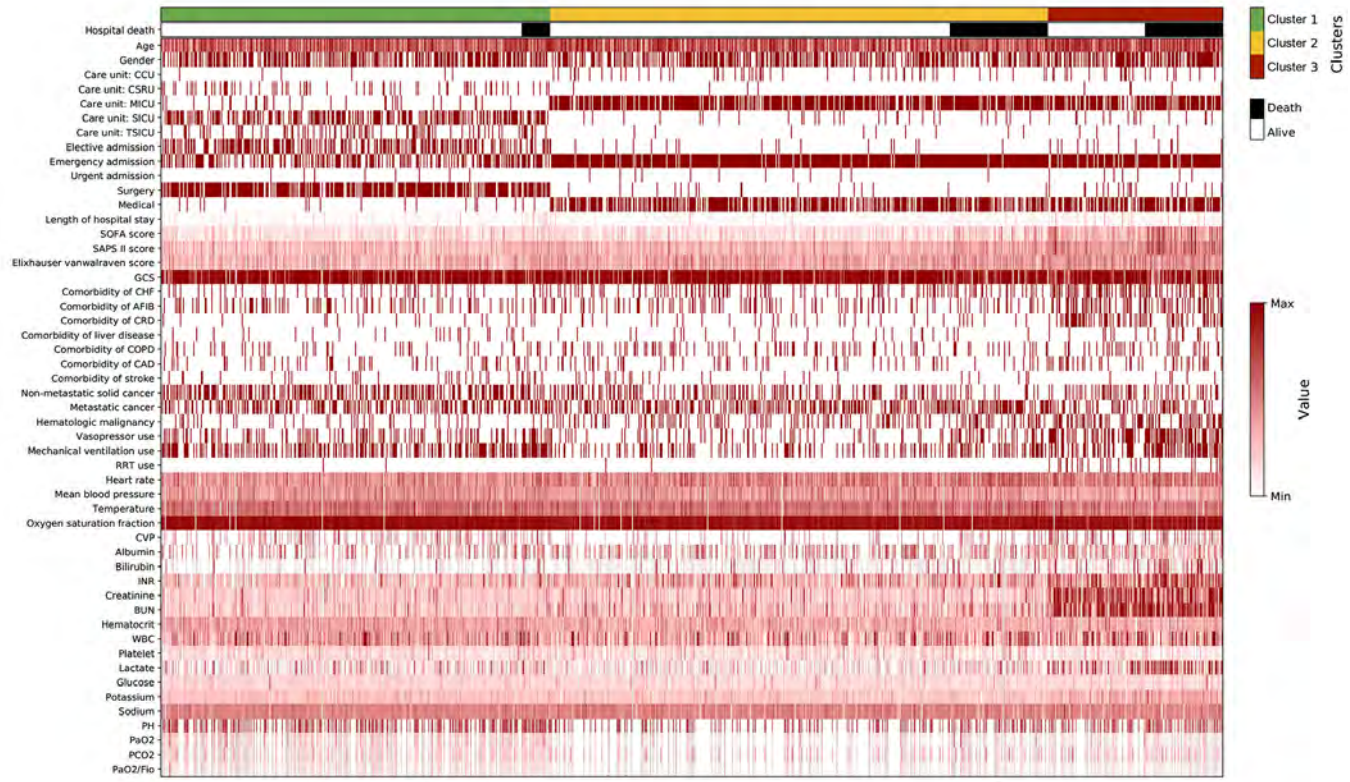

Figure S3 K-means clustering revealed three clinical subtypes. The depth of red color in the heatmap referred to the relative value of the variables. Names of the variables were annotated on the left. CCU, cardiac care unit; CSRU, Cardiac surgery recovery unit; MICU, medical intensive care unit; SICU surgery intensive care unit; TSICU, trauma/surgical intensive care unit; SOFA, sequential organ failure assessment; SAPSII, Simplified Acute Physiology Score II; GCS, Glasgow Coma Scale; CHF, chronic heart failure; AFIB, atrial fibrillation; CRD, chronic renal disease; COPD, chronic obstructive pulmonary disease; CAD, coronary artery disease; RRT, renal replacement treatment; CVP, central venal pressure; INR, international normalized ratio; BUN, blood urea nitrogen; WBC, white blood cells; PaO_2, artery partial pressure of oxygen; PCO_2, partial pressure of carbon dioxide; FiO_2, fraction of inspired oxygen.

http://dx.doi.org/10.21037/atm-20-4634

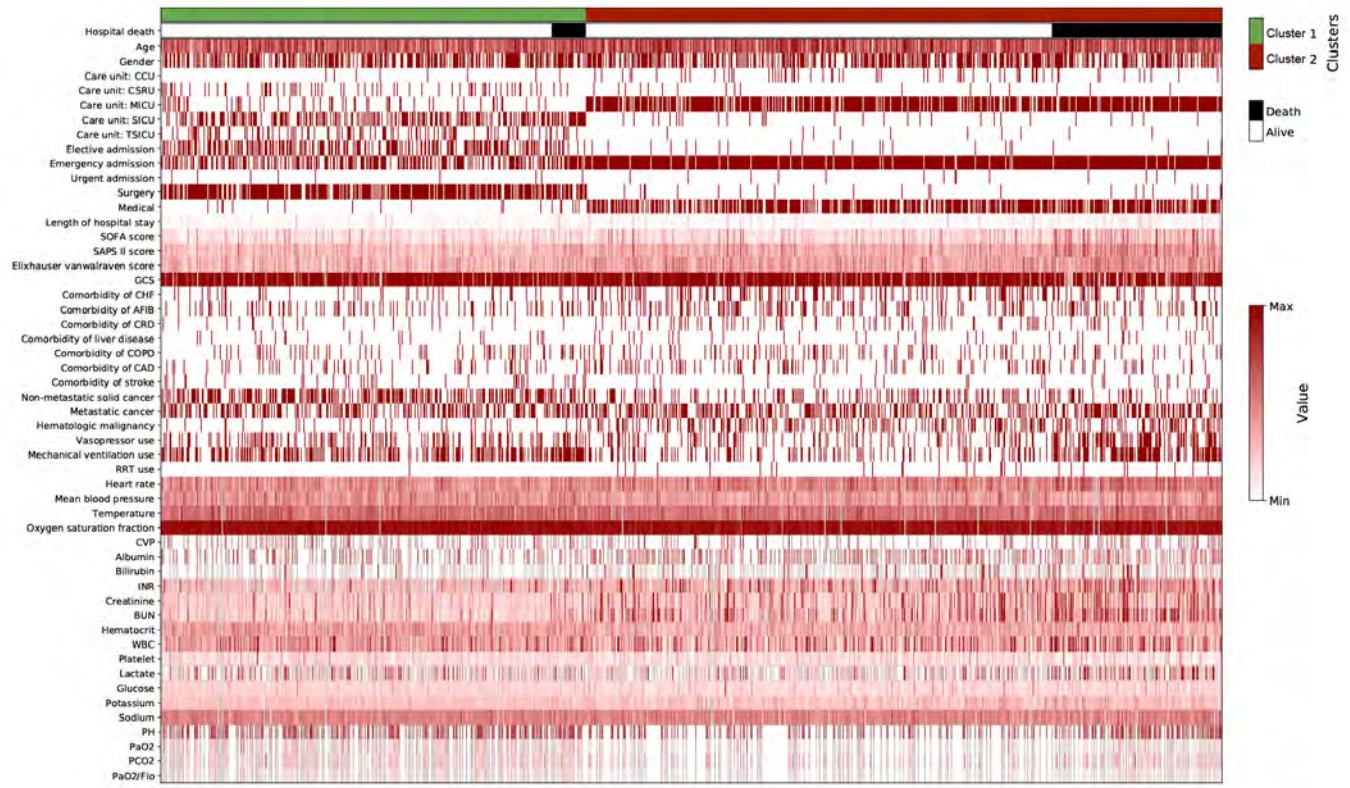

Figure S4 K-means clustering revealed three clinical subtypes. The depth of red color in the heatmap referred to the relative value of the variables. Names of the variables were annotated on the left. CCU, cardiac care unit; CSRU, Cardiac surgery recovery unit; MICU, medical intensive care unit; SICU, surgery intensive care unit; TSICU, trauma/surgical intensive care unit; SOFA, sequential organ failure assessment; SAPSII, Simplified Acute Physiology Score II; GCS, Glasgow Coma Scale; CHF, chronic heart failure; AFIB, atrial fibrillation; CRD, chronic renal disease; COPD, chronic obstructive pulmonary disease; CAD, coronary artery disease; RRT, renal replacement treatment; CVP, central venal pressure; INR, international normalized ratio; BUN, blood urea nitrogen; WBC, white blood cells; PaO_2, artery partial pressure of oxygen; PCO_2, partial pressure of carbon dioxide; FiO_2, fraction of inspired oxygen.

http://dx.doi.org/10.21037/atm-20-4634

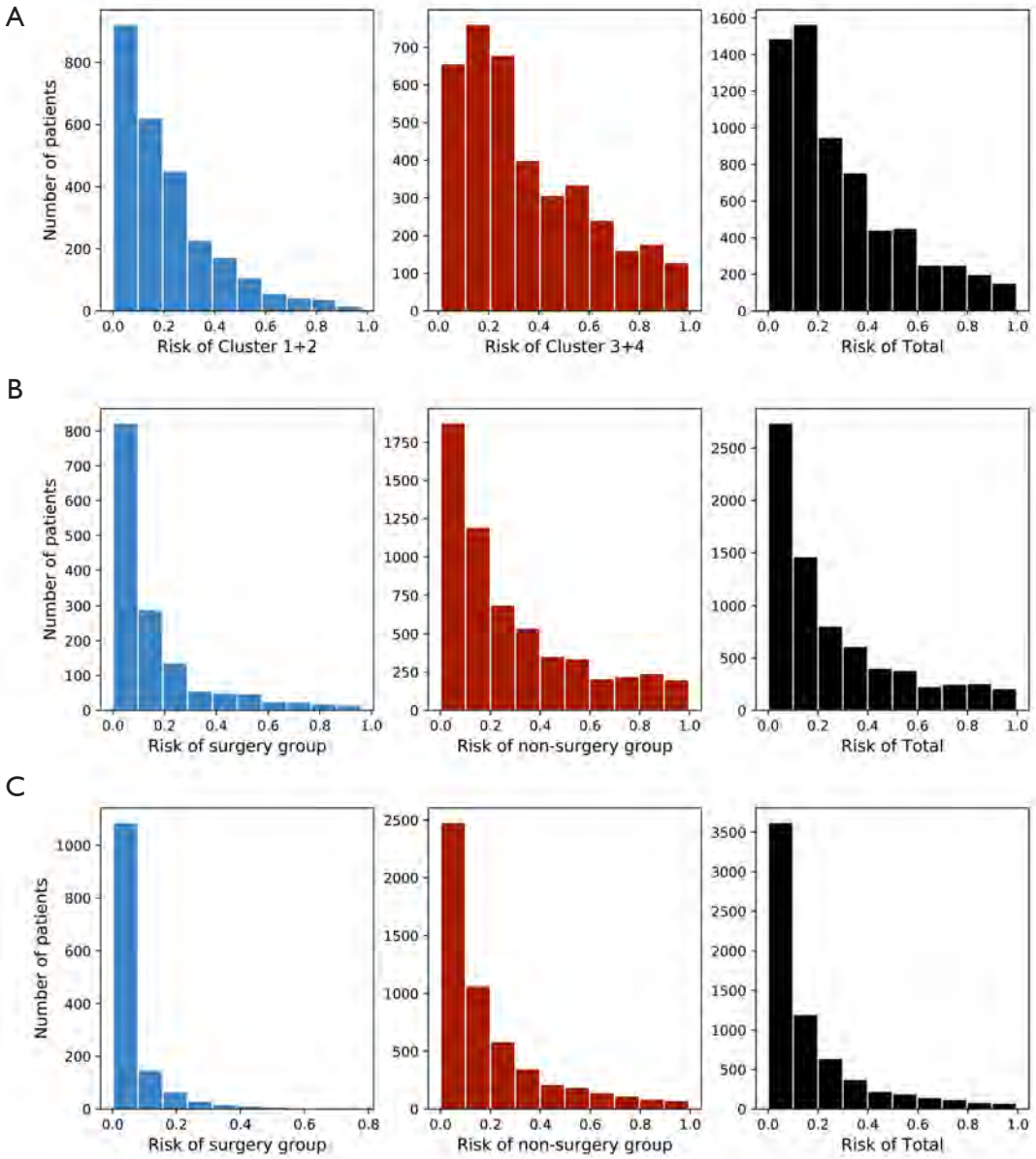

Figure S5 The number of patients in each risk decile of different clusters or groups for (A) SAPS II in MIMIC database; (B) SAPSII in eICU database; (C) APACHE IV in eICU database.

http://dx.doi.org/10.21037/atm-20-4634