

Predictors of All-Cause Mortality among Stroke Patients in Kazakhstan: A Retrospective Study

Student: Artem Yermakov

Supervisor: Abduzhappar Gaipov, MD, PhD

Bachelor of Medical Sciences
Nazarbayev University School of Medicine

Astana, 2025

Abstract

Aim: To examine all-cause mortality predictors among lab indicators and drugs administered to stroke patients in Kazakhstan.

Materials and methods: This retrospective study analyzed data from 272 patient records derived from the UNEHS database (2014-2019). Stroke patients were identified using ICD codes; the GPT-4o model was trained to dissect the records and assist in the extraction of lab and medication data; other clinical and demographic data were derived from the previous study [6] using RPN IDs. Statistical analyses included univariate, multivariate, and imputed multivariate logistic regressions (STATA version 16.1).

Results: In our cohort, deceased patients were older (66.1 vs. 58.6, $p < 0.0001$). Adjusted logistic regression revealed age (OR 1.04), hemoglobin (OR 0.97), and piracetam (OR 3.17) as independent predictors. After multiple imputation, Russian ethnicity (OR 2.79), PTI (OR 0.97), and dopamine (OR 5.23) became independent predictors, while piracetam lost its significance.

Conclusions: Findings suggest the importance of future research on stroke predictors in bigger cohorts to facilitate the introduction of innovations to stroke patients' care to potentially decrease the burden on the Kazakhstani population.

Keywords: Stroke, retrospective cohort study, all-cause mortality, AI-assisted data extraction, predictors of mortality, laboratory data, medication data, regression analysis.

Introduction

Stroke has long been a major public health problem. In 2021, stroke was the third largest cause of death in the world [1]. Kazakhstan is no exception, and the third largest cause of death is stroke as well, taking 224.4 lives per 100,000 population in 2021 [2]. The American Stroke Association has defined stroke as "brain, spinal cord, or retinal cell death resulting from ischemia, based on pathological, imaging, or other objective evidence of cerebral, spinal cord, or retinal focal ischemic injury in a defined vascular distribution; or clinical evidence of cerebral, spinal cord, or retinal focal ischemic injury based on ≥ 24 hours of symptoms or until death, and excluding other causes" [3]. As projected by the Global Burden of Disease (2024), in 2021, a stroke crude incidence was 11.95 million cases [4]. In the same year, the global prevalence of stroke was 93.82 million people, of which acute ischemic stroke accounted for 69.94 million, intracerebral hemorrhage for 16.60 million, and subarachnoid hemorrhage for 7.85 million cases. Looking at mortality, a total of 7.25 million stroke deaths were reported worldwide. Ischemic stroke caused 3.59 million deaths, intracerebral hemorrhage caused 3.31 million deaths, and subarachnoid hemorrhage caused 0.35 million deaths. [5]. In Kazakhstan, the incidence of stroke was valued at 1884 people per million population (PMP) [6].

Early prediction of stroke outcome can help healthcare workers prioritize high-risk patients, optimize treatment regimens, and make better use of resources. This will improve the prognosis of patients, ultimately reducing disability and mortality rates among stroke patients. Different research works show that the utilization of laboratory testing results has great promise to be a reliable tool for that purpose.

Cohen et al. (2015) analyzed lab results of 5,886 acute ischemic stroke patients and reported that admission hemoglobin and sodium values independently increased the probability of patients' successful home discharge [7].

In another study, Carter et al. (2007) investigated predictive variables of all-cause mortality (>30 days after acute event) of acute ischemic stroke patients. The researchers included multiple

variables, including hemostatic as well as biochemical and hematologic variables. Among biochemical variables, lower levels of albumin and elevated levels of creatinine predicted all-cause mortality, whereas among hemostatic variables, higher levels of vWF and β TG were associated with a greater risk of long-term death [8].

Another promising indicator is a neutrophil lymphocyte ratio. Gökhan et al. (2013) enrolled 654 patients with acute ischemic stroke and deduced that the group of deceased patients had a significantly higher neutrophil lymphocyte ratios than the surviving group [9]. Moreover, the researchers found that dead patients had a higher mean neutrophil count than surviving patients [9].

Zhao et al. (2021) examined red blood cell distribution width-albumin ratio (RA) values. The researchers found that higher red blood cell distribution width, lower albumin, and consequently their higher ratio were all independent predictors of stroke patients' mortality [10]. Furthermore, results of this study indicate that several other variables affected RA values and could also potentially be predictors of stroke mortality. Specifically, higher creatinine, lower hematocrit, lower platelet counts, higher international normalized ratio, higher prothrombin time, higher blood urea nitrogen, and higher white blood cell counts all increased the RA values [10]. Although platelet counts did not have a significant p-value in this study, it still should not be ignored.

Additionally, data from retrospective research by Perlstein et al. (2008) suggests that the level of total bilirubin in serum can also show predictive properties with regard to stroke prevalence and outcomes [11]. Specifically, higher serum total bilirubin was associated with a 10% lower probability of adverse outcomes in stroke patients and 9% lower probability of stroke occurrence among all patients [11].

Lastly, Bruno et al. (2002) conducted a study on the impact of admission glucose levels on outcomes of 624 acute ischemic stroke patients [12]. The results suggest that higher blood sugar levels were linked with poorer 3-month outcomes in acute ischemic stroke patients [12].

The majority of the studies listed are biased toward Western populations, therefore, it is important to contribute to global health equality by conducting similar studies on an underrepresented population of Kazakhstan. Unique demographic and genetic factors could influence the results and expand our understanding of the matter. For that reason, this study aims to examine all-cause mortality predictors found in lab indicators of stroke patients in Kazakhstan. By design, this is a retrospective study that utilizes anonymized data from patients' hospital discharge records derived from the Unified Electronic Healthcare System of Kazakhstan, dated between 2014 and 2019. An AI model was trained and used for data extraction from the records.

This study aims examine all-cause mortality predictors among lab indicators and drugs administered to stroke patients in Kazakhstan. We hypothesize that independent predictors of stroke mortality can be found in patients' laboratory and medication data.

Materials and methods

Study design

This research is retrospective by design because it encompasses patients' hospital records obtained from the Unified National Electronic Health System (UNEHS) during the years 2014-2019. Stroke events were identified using ICD codes and demographic data were analyzed by the authors of the original study [6]. The original medical records had personal information of a patient, hospital, and a physician, status, anamnesis, diagnosis, and outcome of a patient. The results of the laboratory tests and both outpatient and in-hospital drug treatment were present. The 272 medical records were analyzed for the laboratory indicators' values and use of drug treatment, including substances, specifically used for the treatment of stroke. The study omitted any personal data.

Data extraction

The utilization of electronic medical records was accomplished with the assistance of the OpenAI GPT4o model in the range of the Grant Program of "Faculty Development Competitive Research (AI and Data Science) 2024-2026". In the scope of this paper, we first trained the AI model to correctly section patients' hospital records. As it can be seen from the Figure 1, to achieve that, we manually annotated roughly 2300 records and divided them in following sections: complaints, anamnesis, status, doctor's visits, lab, instrumental tests, drugs in hospital, procedures in hospital, diagnosis, prescription, recommendation, and outcome. The interface used for that purpose is depicted in Figure 2.

Figure 1

The data extraction process of electronic medical records

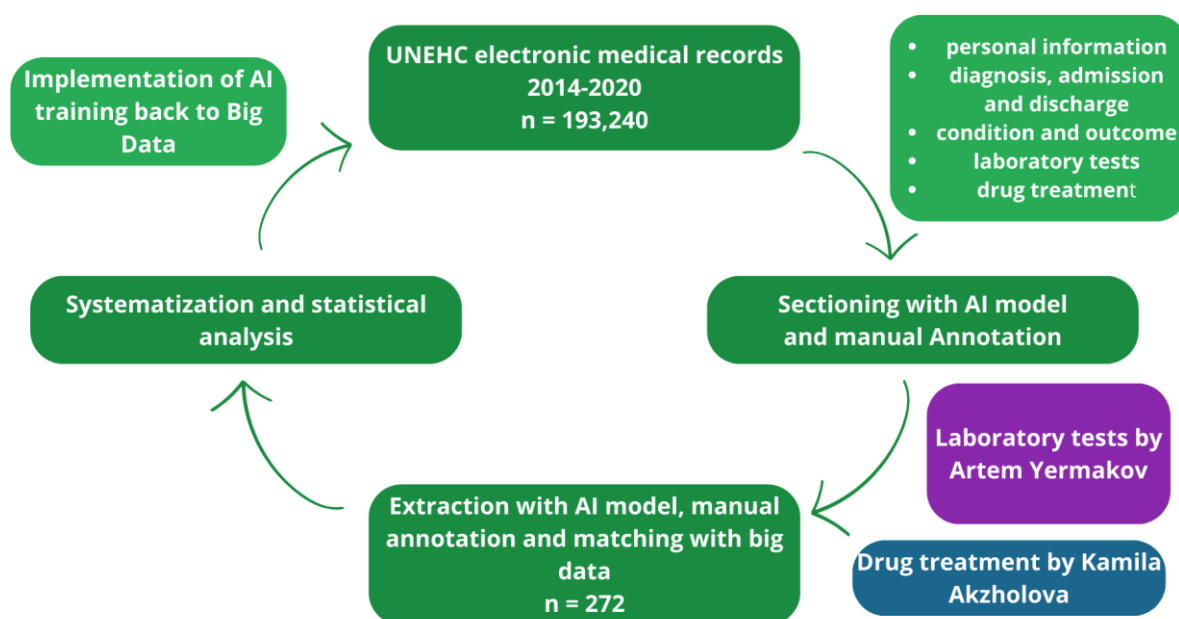


Figure 2

The interface of the annotator for sectioning of the records

Document ID: 663cf14554e3539e72b7aa36

The screenshot shows the interface of the annotator. At the top, there is a document ID: 663cf14554e3539e72b7aa36. Below it, a horizontal bar contains various colored tabs representing different data categories: LAB, DRUGS IN HOSPITAL, PRESCRIPTION, ANAMNESIS, INSTRUMENTAL TESTS, DOCS, STATUS, OUTCOME, PROCEDURES IN HOSPITAL, COMPLAINTS, RECOMMENDATIONS, PERSONAL DATA, and DIAGNOSIS. The main text area displays a medical record snippet. The first line is highlighted in yellow: '1) Внутримозговое кровоизлияние в полушарие кортикальное.' The subsequent text describes the patient's condition and treatment, with several lines highlighted in yellow and blue. The text includes details about the patient's state, vital signs, and specific medical interventions, such as 'Вит. «С» - 6,0 мл + физ. раствор 200,0 в/в; Этамзилат 12,5% - 2,0 x 2 р/д в/м; Фуросемид 20 мг в/в; геллофузин 500 мл в/в; цефазолин 1,0 x 4 р/д в/в; дифосфацин 1000 мг + физ. раствор 200,0 в/в; зинап 10 мг x 3 р/д; кордафен 10 мг - 2 р/д; конвулекс 500 мг x 2 р/д; церулин - 2,0 в/в; зинап 1,0 + физ. раствор 200,0 в/в; эбрантил 75 мг + физ. раствор 200,0 в/в; MgSO4 - 25% - 2,0 в/в; глюкоза 10% - 400 мл + инсулин 14 ЕД в/в; цефотаксим 1,0 x 4р/д; каптоспес 50 мг 1 р/д. КТ головного мозга от 15.12.2013 г.: КТ - признаки геморрагического инсульта в бассейне СМА слева, гематома, с прорывом в желудочковую систему, энцефалопатии, отека головного мозга. Постинсультные изменения в бассейне ЗМА слева. УЗИ почек от 20.12.13 - Заключение: воспалительные изменения в почках.'

Based on that, the AI model was trained using Large Language Model (LLM) learning. In the next step, we used the trained model to extract drug and lab data and manually verified all the values. The interface used for drug data extraction is shown in Figure 3. As a result, we were able to obtain laboratory and medication data for 272 patients.

Figure 3

The interface of the training model for the extraction of drug treatment data

Raw Drugs Text

Raw drugs text

этамзилат, трамадол, димедрол, кетотоп, аскорутин, MgSO4, пирацетам, нимотоп, маннит, вит. С, церулин, КМА, глюкоза-инсулин, КС, цефазолин, эгилон, зиналприл, коринин, актовегин, трисоль,

extracted data from electronic medical records:
here prescription list of a patient

Structured Drugs Data

e.g. tablet, solution, etc amount of the drug e.g. once a day, twice a day e.g. intravenous

Name	Form	Dosage	Unit	Concentration	Frequency	Duration	Route	Delete
Этамзилат	NA	NA		NA	NA	NA	NA	
Трамадол	NA	NA		NA	NA	NA	NA	
Димедрол	NA	NA		NA	NA	NA	NA	
Кетотоп	NA	NA		NA	NA	NA	NA	
Аскорутин	NA	NA		NA	NA	NA	NA	
MgSO4	NA	NA		NA	NA	NA	NA	
Пирацетам	NA	NA		NA	NA	NA	NA	

The data for 272 were then merged with a big dataset from the original study, using RPN IDs [6]. This was done to address the limitation of the aforementioned study, which is the absence of data on laboratory tests and drug treatment of patients in the analysis.

Ethical statement

Nazarbayev University IREC granted approval to the studies involving human subjects (NU-IREC 490/18112021). The research did not violate any institutional requirements or legal legislations. The requirement for informed consent was waived since the study is retrospective and utilizes depersonalized data. All personal data including names of patients, doctors, hospitals, and cities were deleted using the AI model and were never included in the database, nor mentioned in the report.

Demographic characteristics

Unique patient entries consisted of gender, date of birth, ethnicity, living area (urban, rural), type of admission, and death date if they were applicable.

Medical characteristics

Types of stroke (AIS, ICH, SAH), comorbidities, follow-up period, length of hospital stay, acute kidney injury (AKI), all-cause, and in-hospital mortalities were also present where applicable. The comorbidities included diabetes and hypertension. AKI was determined using the difference of more than 26 $\mu\text{mol/l}$ between first and second creatinine measurements, where applicable, in accordance with the National Kidney Foundation [13]. The death during the hospitalization period was categorized as in-hospital mortality, while deaths during the whole observation period were accounted for as all-cause mortality.

Drug treatment

From the extracted data with AI, overall 173 unique drugs were identified. The 273 medical records were categorized for the use of a list of 11 drugs used in the stroke treatment, using a binary system (yes/ no). The 11 drugs: captopril, diazepam, dexamethasone, dopamine, magnesium sulphate, fraxiparin, actovegin, cerebrolysin, vinposan, piracetam, strocitus were chosen on the basis of clinical protocol of Ministry of Health and Social Development of the Republic of Kazakhstan dated July 13, 2016, Protocol No. 7 with the code I 64 - stroke, not specified as hemorrhage or infarction.

Laboratory tests

Laboratory indicators included are: creatinine, total serum bilirubin, fibrinogen, blood leukocyte count, hemoglobin, hematocrit, blood sugar, blood urea, erythrocyte sedimentation rate (ESR), potassium, sodium, blood protein, urea protein, alanine aminotransferase (ALT), aspartate transaminase (AST), prothrombin time (PTT), activated partial thrombin time (APTT), international normalized ratio (INR), and prothrombin index (PTI). AI was used to isolate sections with lab results from the medical records, while the extraction and processing of numeric data from these sections was done manually.

Statistical analysis

Descriptive data are depicted as percentages for categorical variables and as mean $\pm$ standard deviation (SD) in case of normal distribution or median (interquartile range—IQR) in other cases. All subjects were divided into survivors and deceased. χ^2 test was utilized for comparison of categorical variables, while continuous variables were compared using Student's t tests or Mann–Whitney U tests, as appropriate. The association between demographics, laboratory data, medication data, and risk of all-cause mortality was examined using unadjusted (univariate) and adjusted (multivariate) logistic regression analyses. The following models were constructed that include confounders picked based on their significance level in unadjusted analysis and overall potential health effect: Model 1: age, gender, ethnicity, length of hospital stay; Model 2: added Dopamine, Fraxiparine, Piracetam, and Strocit to the model 1; Model 3: added fibrinogen, hemoglobin, INR, PTI, blood sugar, and urea to the model 2. Subsequent logistic regression with multiple imputation was done as a sensitivity analysis to account for missing values.

All P values displayed are two-sided with level of significance set at <0.05 for all analyses. All analyses were conducted using STATA 16.1 (College Station, TX).

Results

Baseline characteristics

A total of 272 patients were included in the analysis. Among them, 126 (46.3%) died, and 146 (53.7%) survived during the follow-up period. The baseline demographic and clinical characteristics are shown in Table 1.

Table 1

General characteristics of admitted stroke patients.

Characteristic	Total (n = 272)	Survived (n = 146)	Deceased (n = 126)	p-value
Age, year (mean ± SD)	62.1±13.61	58.6±13.31	66.1±12.87	<0.0001
Gender, n (%)				0.380
Male	133 (48.90)	75 (51.37)	58 (46.03)	
Female	139 (51.10)	71 (48.63)	68 (53.97)	
Ethnicity, n (%)				<0.0001
Kazakh	146 (53.68)	91 (62.33)	55 (43.65)	
Russian	75 (27.57)	25 (17.12)	50 (39.68)	
Other	51 (18.75)	30 (20.55)	21 (16.67)	
Laboratory findings at admission				
Serum creatinine, μmol/l (median [IQR])	87.67 [68.3-113]	84 [62.31-105.7]	90 [75.9-120.6]	0.070
ALT, U (median [IQR])	16 [10-29.87]	16 [10-32.19]	14.9 [9.6-28]	0.453
aPTT, seconds (mean ± SD)	35.69±14.51	35.87±13.41	35.48±15.85	0.872
AST, U (median [IQR])	19.9 [10.8-31.8]	23.15 [11.4-32.2]	18.33 [10.8-29.9]	0.277
Total bilirubin, μmol/l (median [IQR])	13.68 [9.8-18.2]	13.9 [10-18]	13.2 [9.8-18.5]	0.582
Fibrinogen, g/l (mean ± SD)	3.81±1.65	3.60±1.42	4.09±1.9	0.077
Hemoglobin, g/l (mean ± SD)	136.19±21.57	139.04±19.46	132.58±23.61	0.029
Hematocrit, % (mean ± SD)	42.28±8.78	43.19±6.59	41.39±10.49	0.273
Leukocytes, *10 ⁹ /l (mean ± SD)	8.81±3.92	8.68±3.68	8.98±4.22	0.572
INR, (mean ± SD)	1.07±0.19	1.05±0.15	1.10±0.23	0.140
Potassium, mmol/l (mean ± SD)	4.07±0.86	4.13±0.92	3.99±0.80	0.423
Blood protein, g/l (mean ± SD)	67.95±10.81	68.48±10.26	67.28±10.13	0.472
PTI, % (mean ± SD)	91.87±10.43	94.29±8.46	89.22±11.73	0.0049
PTT, seconds (median [IQR])	15.5 [14-17.8]	15.2 [14-16.5]	16 [14.3-18.55]	0.118
Sodium, mmol/l (mean ± SD)	141.63±7.68	141.40±6.67	141.90±8.81	0.772
ESR, mm/h (median [IQR])	13 [6-23]	13 [5-23]	12.5 [6-23]	0.592
Blood sugar, mmol/l (median [IQR])	6.2 [5.3-7.85]	6.2 [5.3-7.9]	6.2 [5.3-7.7]	0.888
Urine protein, g/l (median [IQR])	0.03 [0-0.066]	0.033 [0-0.066]	0 [0-0.066]	0.939
Urea, mmol/l (median [IQR])	6.1 [4.6-7.6]	5.8 [4.4-6.9]	6.37 [5.2-8.82]	0.0044
Medications				
Captopril, n (%)	3 (1.10)	3 (2.05)	0 (0.00)	0.106
Diazepam, n (%)	3 (1.10)	0 (0.00)	3 (2.38)	0.061
Dexamethasone, n (%)	27 (9.93)	14 (9.59)	13 (10.32)	0.841
Dopamine, n (%)	14 (5.15)	3 (2.05)	11 (8.73)	0.013
Magnesium sulfate, n (%)	155 (56.99)	87 (59.59)	68 (53.97)	0.350

Fraxiparine (Nadroparin), n (%)	19 (6.99)	15 (10.27)	4 (3.17)	0.022
Actovegin, n (%)	141 (51.84)	74 (50.68)	67 (53.17)	0.682
Cerebrolysin, n (%)	32 (11.76)	19 (13.01)	13 (10.32)	0.491
Vinposan (Vinpocetine), n (%)	61 (22.43)	31 (21.23)	30 (23.81)	0.611
Piracetam, n (%)	133 (48.90)	64 (43.84)	69 (54.76)	0.072
Strocit (Citicoline), n (%)	62 (22.79)	39 (26.71)	23 (18.25)	0.097
Complication/outcome				
Length of hospital stay, days (median [IQR])	11.5 [9-15]	12 [10-16]	11 [8-14]	0.042
Follow up period, days (median [IQR])	2568 [330.5-2579]	2579[2572-2584]	187.5 [12-1086]	<0.0001
AKI, n (%)	12 (4.41)	5 (3.42)	7 (5.56)	0.393
Type of stroke, n (%)				0.965
AIS	118 (43.48)	63 (43.15)	55 (43.65)	
ICH	144 (52.94)	78 (53.42)	66 (52.38)	
SAH	10 (3.68)	5 (3.42)	5 (3.97)	
Type of death, n (%)				<0.0001
Early	44 (16.18)	0 (0.00)	44 (34.92)	
Late (>30 days)	82 (30.15)	0 (0.00)	82 (65.08)	

CKD, chronic kidney disease; ALT, alanine aminotransferase; APTT, activated partial thrombin time; AST, aspartate transaminase; INR, international normalized ratio; PTI, prothrombin index; PTT, prothrombin time; ESR, erythrocyte sedimentation rate; AKI, acute kidney injury; AIS, acute ischemic stroke; ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.

The mean age of the cohort was 62.1 ± 13.61 years, with non-survivors being older on average (66.1 ± 12.87) compared to survivors (58.6 ± 13.31). The gender distribution was 48.9% male and 51.1% female, with no significant difference between groups ($p = 0.38$). In terms of ethnicity, 53.68% were Kazakh, 27.57% Russian, and 18.75% other ethnicities, with significant difference in mortality between these groups ($p < 0.0001$). Kazakhs were more likely to survive (62.33% survived), the opposite is true for Russians (33.3% survived), while among other ethnicities there are a little bit more survived than diseased patients (58.8% survived). For laboratory indicators deceased patients had a lower mean hemoglobin, lower mean prothrombin index, and higher median urea level compared to survivors. With regard to drugs, deceased patients had a higher administration of dopamine and lower administration of fraxiparine.

Median length of hospitalization was 11.5 days and survived patients stayed in hospital for a longer time than non-survivors. The majority of patients (52.94%) was diagnosed with intracerebral hemorrhage, 43.48% of cases accounted for acute ischemic stroke, while

subarachnoid hemorrhage accounted for remaining 3.68%. Early death was less common among deceased patients accounting for 34.92% of cases.

Association between demographic and clinical data with survival

The univariate logistic regression was used to characterize associations between clinical-demographic parameters and death risk is summarized in Table 2. Each increase by 5 years of age resulted in 26% higher risk of death (Odds Ratio – OR 1.26; 95% Confidence Interval – CI 1.14-1.40; p-value (p) < 0.0001). People of Russian ethnicity had almost 3-fold greater odds of death (OR 2.86; 95% CI 1.37-5.96; p = 0.005) compared to other nationalities (not Russian, nor Kazakh). Each incremental decrease by 10 g/l of hemoglobin is associated with 15% higher risk of death (OR 1.15; 95% CI 1.01-1.31; p = 0.031). Each decrease of PTI by 10% increased odds of death by 37% (OR 1.37; 95% CI 1.01-1.84; p = 0.043). Urea showed 2-fold increase in death odds with each elevation by 5 mmol/l (OR 2.01; 95% CI 1.25-3.23; p = 0.004). Administration of dopamine demonstrated 4.5-fold increase in odds of death (OR 4.56; 95% CI 1.24-16.73; p = 0.022). Application of fraxiparine resulted in 71% lower risk of fatal outcome (OR 0.29; 95% CI 0.09-0.89; p = 0.030). Increase of follow up period is connected to 9% lower odds of death (OR 0.91; 95% CI 0.86-0.96; p = 0.001).

Table 2

The association between demographics, medication, laboratory related parameters and risk of all-cause mortality using univariate logistic regression models.

Characteristic	Number of observations / Number of events	Univariate OR (95% CI)	p-value
Age	272 / 126	1.05 (1.03-1.07)	<0.0001
Age +5	272 / 126	1.26 (1.14-1.40)	<0.0001
Gender (female ref.)			
Male	133 / 58	0.81 (0.50-1.30)	0.380
Ethnicity (other ref.)			
Kazakh	146 / 55	0.86 (0.45-1.66)	0.658
Russian	75 / 50	2.86 (1.37-5.96)	0.005
Serum creatinine	189 / 89	1.00 (1.00-1.01)	0.107
ALT	154 / 75	0.99 (0.97-1.01)	0.258
aPTT	146 / 66	0.998 (0.98-1.02)	0.871
AST	146 / 70	0.99 (0.98-1.01)	0.336

Total bilirubin	183 / 82	0.997 (0.97-1.02)	0.835
Fibrinogen	145 / 62	1.21 (0.97-1.49)	0.088
Hemoglobin	216 / 95	0.99 (0.97-1.00)	0.031
Hemoglobin -5	216 / 95	1.07 (1.01-1.15)	
Hemoglobin -10	216 / 95	1.15 (1.01-1.31)	
Hematocrit	115 / 58	0.98 (0.93-1.02)	0.280
Leukocytes	216 / 95	1.02 (0.95-1.09)	0.570
INR	148 / 62	3.80 (0.62-23.39)	0.150
Potassium	87 / 40	0.81 (0.49-1.35)	0.420
Blood protein	153 / 68	0.99 (0.99-1.02)	0.472
PTI	134 / 64	0.97 (0.94-1.00)	0.043
PTI -5	134 / 64	1.17 (1.01-1.36)	
PTI -10	134 / 64	1.37 (1.01-1.84)	
PTT	105 / 44	1.00 (0.97-1.03)	0.882
Sodium	81 / 37	1.01 (0.95-1.07)	0.769
ESR	181 / 82	1.01 (0.99-1.03)	0.335
Blood sugar	208 / 93	1.05 (0.98-1.11)	0.175
Urine protein	192 / 84	2.14 (0.57-7.95)	0.257
Urea	186 / 85	1.15 (1.05-1.26)	0.004
Urea +5	186 / 85	2.01 (1.25-3.23)	
Dexamethasone	272 / 126	1.09 (0.49-2.40)	0.841
Dopamine	272 / 126	4.56 (1.24-16.73)	0.022
Magnesium sulfate	272 / 126	0.80 (0.49-1.29)	0.351
Fraxiparine	272 / 126	0.29 (0.09-0.89)	0.030
Actovegin	272 / 126	1.10 (0.69-1.78)	0.682
Cerebrolysin	272 / 126	0.77 (0.36-1.63)	0.492
Vinposan	272 / 126	1.16 (0.66-2.05)	0.612
Piracetam	272 / 126	1.55 (0.96-2.51)	0.073
Strocut	272 / 126	0.61 (0.34-1.10)	0.099
Length of hospital stay	272 / 126	0.96 (0.93-1.00)	0.070
Follow up period	272 / 126	0.91 (0.86-0.96)	0.001
AKI	272 / 126	1.66 (0.51-5.36)	0.398
Stroke type (AIS ref.)			
ICH	144 / 66	0.97 (0.60-1.58)	0.900
SAH	10 / 5	1.15 (0.32-4.17)	0.837

Captopril and diazepam omitted due to collinearity. CKD, chronic kidney disease; ALT, alanine aminotransferase; APTT, activated partial thrombin time; AST, aspartate transaminase; INR, international normalized ratio; PTI, prothrombin index; PTT, prothrombin time; ESR, erythrocyte sedimentation rate; AKI, acute kidney injury; AIS, acute ischemic stroke; ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.

Adjustment using different models is depicted in Table 3. Age (OR 1.04; 95% CI 1.01-1.08; p = 0.023), hemoglobin level (OR 0.97; 95% CI 0.94-0.99; p = 0.008), and administration of piracetam (OR 3.17; 95% CI 1.20-8.41; p = 0.020). Dopamine showed collinearity.

Table 3

The association between demographics, medication, laboratory related parameters and risk of all-cause mortality using multivariate logistic regression models.

Covariates	Model 1 OR (95% CI)	p-value	Model 2 OR (95% CI)	p-value	Model 3 OR (95% CI)	p-value
Age	1.04 (1.02-1.06)	<0.0001	1.04 (1.01-1.08)	0.017	1.04 (1.01-1.08)	0.023
Gender (female ref.)	0.99 (0.59-1.65)	0.958	1.16 (0.50-2.73)	0.726	1.07 (0.42-2.74)	0.891
Ethnicity (other ref.)						
Kazakh	1.07 (0.54-2.10)	0.854	1.57 (0.54-4.55)	0.405	1.99 (0.61-6.44)	0.253
Russian	2.73 (1.28-5.81)	0.009	4.05 (1.18-13.86)	0.026	2.79 (0.74-10.52)	0.129
Hemoglobin	-	-	0.97 (0.95-0.99)	0.011	0.97 (0.94-0.99)	0.008
PTI	-	-	0.98 (0.95-1.01)	0.216	0.98 (0.94-1.01)	0.249
Urea	-	-	1.08 (0.97-1.20)	0.171	1.10 (0.96-1.27)	0.163
Dopamine	-	-	-		1	-
Fraxiparine	-	-	-		0.20 (0.03-1.28)	0.088
Piracetam	-	-	-		3.17 (1.20-8.41)	0.020
Strocit	-	-	-		0.38 (0.11-1.28)	0.119

Dopamine omitted due to collinearity. INR, international normalized ratio; PTI, prothrombin index.

Table 4 illustrates the same adjusted models after multiple imputation were performed. Age (OR 1.04; 95% CI 1.02-1.07; $p = 0.001$) and hemoglobin level (OR 0.98; 95% CI 0.96-1.00; $p = 0.045$) remained independent predictors of mortality. Moreover, other parameters became significant predictors. Namely, Russian ethnicity (OR 2.84; 95% CI 1.19-6.79; $p = 0.019$) as well as PTI (OR 0.97; 95% CI 0.94-1.00; $p = 0.043$), and administration of dopamine (OR 5.23; 95% CI 1.06-25.74; $p = 0.042$). At the same time piracetam was no longer significant in the model.

Table 4

The association between demographics, medication, laboratory related parameters and risk of all-cause mortality using multivariate logistic regression models with multiple imputation.

Covariates	Model 1 OR (95% CI)	p-value	Model 2 OR (95% CI)	p-value	Model 3 OR (95% CI)	p-value
Age	1.04 (1.02-1.06)	<0.0001	1.04 (1.02-1.07)	0.001	1.04 (1.02-1.07)	0.001
Gender (female ref.)	0.99 (0.59-1.65)	0.958	1.21 (0.66-2.21)	0.541	1.16 (0.63-2.14)	0.637
Ethnicity (other ref.)						
Kazakh	1.07 (0.54-2.10)	0.854	1.01 (0.47-2.15)	0.986	0.99 (0.45-2.15)	0.977
Russian	2.73 (1.28-5.81)	0.009	2.90 (1.23-6.83)	0.015	2.84 (1.19-6.79)	0.019
Hemoglobin	-	-	0.98 (0.97-1.00)	0.047	0.98 (0.96-1.00)	0.045
PTI	-	-	0.97 (0.94-1.00)	0.041	0.97 (0.94-1.00)	0.043
Urea	-	-	1.09 (0.99-1.20)	0.065	1.09 (0.98-1.21)	0.104
Dopamine	-	-	-		5.23 (1.06-25.74)	0.042
Fraxiparine	-	-	-		0.25 (0.07-1.01)	0.052
Piracetam	-	-	-		1.67 (0.94-2.97)	0.082
Strocit	-	-	-		0.83 (0.41-1.68)	0.600

INR, international normalized ratio; PTI, prothrombin index.

Discussion

Our results suggest that patient's age is connected to elevation of odds of death in all regressions used. This result aligns with findings of the previous study [6]. Higher incidence of complications and higher mean number of chronic conditions in older people are possible explanations to this outcome [14].

Unadjusted logistic regression and adjusted model after imputation show that people of Russian ethnicity have a substantially higher risk of death compared to Kazakhs and people of other ethnicities. Similar results are demonstrated in the original study [6], however odds ratio in our work is much higher. This discrepancy can potentially be attributed to some cultural or genetic differences. Maybe people of Russian descent do not check their health as much as others and therefore have more serious and numerous comorbidities. On top of that, after performing Student's t-test we found out that in this study Russian population was older on average than Kazakh (66.9 vs. 58.8; $t = 4.33$; $p = < 0.0001$). Unfortunately, there are no appropriate studies we could find. Furthermore, our cohort has significantly more Kazakh population, and this ethnical disbalance is a limitation to this outcome.

We found that a higher blood urea level is a predictor of mortality in our unadjusted logistic regression. Previous studies mainly focused on blood urea nitrogen (BUN) or the BUN to creatinine ratio [15,16]. BUN is a component of urea that we have in our study, thus, these measurements are equivalent [17]. In case of our cohort, higher blood urea level is connected to higher odds of death, which aligns with findings of the aforementioned studies. This could be explained by the fact that high blood urea level is a crucial indicator of kidney damage. Therefore, patients with higher levels of urea in their blood were likely to have renal dysfunction, which would substantially increase the risk of death [18].

In previous research, it was found that higher admission levels of hemoglobin increased the odds of stroke patients' survival [7]. Our results support this claim, since lower levels of hemoglobin showed higher odds of fatality in all analyses. Anemia (low hemoglobin) was found

to be associated with mortality in stroke patients in the research by Desai et al. [19]. On top of that, lower hemoglobin is an indicator of iron deficiency, which also contributes to higher odds of stroke mortality [20]. The possible explanations for that is a lowered ability of blood to supply organs, including the brain, with oxygen, which could reduce the chances of patients' survival during ischemic stroke.

Prothrombin index (PTI) is a mathematical ratio of reference thrombin time and a patient's thrombin time multiplied by 100% [21]. It is an indicator of coagulation, and a lower PTI value constitutes lower coagulation. This research suggests that lower PTI is a predictor of stroke mortality both in unadjusted and imputed multivariate logistic regressions. This claim is backed up by previous research on anticoagulants' property to increase the odds of death in cerebral hemorrhage [22]. The majority of our cohort consisted of patients with intracerebral hemorrhage, and a lower ability of blood to coagulate would make this condition more severe due to higher volumes of blood entering the intracerebral space. This is a possible explanation of PTI significance as a parameter of unadjusted regression in this study.

This research identified that administration of dopamine was a significant indicator of a higher risk of death in unadjusted logistic regression and adjusted regression after imputation, although the 95% confidence interval (CI) is very large. Dopamine is used as a peripheral vasostimulant. Low infusion rates (0.5-2 $\mu\text{g/kg/min}$) promote visceral vasodilation and can be used during renal failure to increase urinary flow. Intermediate infusion rates (2-10 $\mu\text{g/kg/min}$) elevate cardiac output via stimulation of electrical conductivity and myocardial contractility. Higher doses cause vasoconstriction and increased blood pressure [23]. Consequently, dopamine is often used in critical care, where patients have lower chances of survival. This is the most probable explanation of the results in our study; however, it could also be due to chance, taking into consideration the large 95% CI.

Another parameter that was significant in the univariate regression is administration of fraxiparine. Fraxiparine (nadroparin) is an anticoagulant sometimes used post-stroke for

prophylaxis of deep vein thrombosis (DVT) [24]. Thus, fraxiparine was often administered to surviving patients to prevent complications, which could explain the lower odds of death.

Piracetam turned out to be a significant predictor of stroke mortality only in adjusted logistic regression, but lost its significance after imputation. Manufacturers claim that this nootropic drug has neuroprotective properties and can ease cognitive recovery after stroke, although existing literature does not provide sufficient evidence regarding the effect of piracetam on stroke rehabilitation [25]. The large 95% CI in adjusted regression suggests that this result could appear due to chance.

This study is primarily limited by a small sample size, which has caused imbalances in ethnic, gender, and stroke type representation in our cohort. Furthermore, the OpenAI model used for data extraction may have made errors, and manual evaluation of extracted data could have suffered from inconsistencies in parameter names and units of measurement. Another limitation arises from laboratory results, including a lot of missing values, although we addressed it using multiple imputation. Finally, the original study by Gulnur Zhakhina, et al. (2022) included various surgery procedures and several comorbidities, unfortunately, we were not able to match these parameters with our cohort due to inability to quickly retrieve these data and time constraints.

Conclusion

In conclusion, this research retrospectively assessed records of 272 stroke patients utilizing AI-assisted data extraction and data from the previous study [6] to understand how laboratory and medication variables are associated with mortality. Age and hemoglobin level were only significant predictors of mortality in the adjusted logistic regression, although PTI also became significant after multiple imputation. These findings have the potential to become important for stroke prognosis and risk assessment in the future. Future research with a bigger population is required to expand understanding of the subject.

References

1. World Health Organization. (2024). *The top 10 causes of death*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
2. World Health Organization. *Kazakhstan data*. World Health Organization. <https://data.who.int/countries/398>
3. Sacco, R. L., Kasner, S. E., Broderick, J. P., Caplan, L. R., Connors, J. J. B., Culebras, A., Elkind, M. S. V., George, M. G., Hamdan, A. D., Higashida, R. T., Hoh, B. L., Janis, L. S., Kase, C. S., Kleindorfer, D. O., Lee, J.-M., Moseley, M. E., Peterson, E. D., Turan, T. N., Valderrama, A. L., & Vinters, H. V. (2013). An updated definition of stroke for the 21st century: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 44(7), 2064–2089. <https://doi.org/10.1161/STR.0b013e318296aeca>
4. GBD 2021 Stroke Risk Factor Collaborators. (2024). Global, regional, and national burden of stroke and its risk factors, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Neurology*, 23(10), 973–1003. [https://doi.org/10.1016/s1474-4422\(24\)00369-7](https://doi.org/10.1016/s1474-4422(24)00369-7)
5. American Heart Association. (2025). *2025 Heart Disease & Stroke Statistical Update Fact Sheet Global Burden of Disease*. American Heart Association. <https://professional.heart.org/-/media/phd-files-2/science-news/2/2025-heart-and-stroke-stat-update/factsheets/2025-stats-update-fact-sheet-global-burden-of-disease.pdf>
6. Gulnur Zhakhina, Bakhytbek Zhalmagambetov, Arnur Gusmanov, Yesbolat Sakko, Sauran Yerdessov, Elzar Matmusaeva, Aliya Imanova, Crape, B., Sarria-Santamera, A., & Abduzhappar Gaipov. (2022). Incidence and mortality rates of strokes in Kazakhstan in 2014–2019. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-20302-8>

7. Cohen, A. S., Siddiqui, K. A., Ali, S. F., Anderson, C. D., Silverman, S. B., & Schwamm, L. H. (2015). Abstract T P267: Admission Hemoglobin and Sodium Values are Important Factors Associated with Good Outcome After Acute Ischemic Stroke. *Stroke*, 46(1).
https://doi.org/10.1161/str.46.suppl_1.tp267
8. Carter, A. M., Catto, A. J., Mansfield, M. W., Bamford, J. M., & Grant, P. J. (2007). Predictive Variables for Mortality After Acute Ischemic Stroke. *Stroke*, 38(6), 1873–1880. <https://doi.org/10.1161/strokeaha.106.474569>
9. Gökhan, S., Özhasenekler, A., Mansur Durgun, H., Akil, E., Üstündag, M., & Orak, M. (2013). Neutrophil lymphocyte ratios in stroke subtypes and transient ischemic attack. *European Review*. <https://www.europeanreview.org/article/3437>
10. Zhao, N., Hu, W., Wu, Z., Wu, X., Li, W., Wang, Y., Zhao, H. (2021). The Red Blood Cell Distribution Width–Albumin Ratio: A Promising Predictor of Mortality in Stroke Patients. *International Journal of General Medicine*.
<https://www.semanticscholar.org/paper/The-Red-Blood-Cell-Distribution-Width%E2%80%93Albumin-A-of-Zhao-Hu/27841bec805414013f82b77494ea79ff4d14c78d>
11. Perlstein, T. S., Pande, R. L., Creager, M. A., Weuve, J., & Beckman, J. A. (2008). Serum Total Bilirubin Level, Prevalent Stroke, and Stroke Outcomes: NHANES 1999–2004. *The American Journal of Medicine*, 121(9), 781-788.e1.
<https://doi.org/10.1016/j.amjmed.2008.03.045>
12. Bruno, A., Levine, S. R., Frankel, M. R., Brott, T. G., Lin, Y., Tilley, B. C., Lyden, P. D., Broderick, J. P., Kwiatkowski, T. G., & Fineberg, S. E. (2002). Admission glucose level and clinical outcomes in the NINDS rt-PA Stroke Trial. *Neurology*, 59(5), 669–674.
<https://doi.org/10.1212/wnl.59.5.669>
13. National Kidney Foundation. (2024). *Acute kidney injury (AKI)*. National Kidney Foundation. <https://www.kidney.org/kidney-topics/acute-kidney-injury-aki>

14. Chen, C.-M., Hsu, H.-C., Chang, C.-H., Lin, C.-H., Chen, K.-H., Hsieh, W.-C., & Chang, W.-M. (2014). Age-based prediction of incidence of complications during inpatient stroke rehabilitation: a retrospective longitudinal cohort study. *BMC Geriatrics*, 14(1).
<https://doi.org/10.1186/1471-2318-14-41>
15. Schrock, J. W., Glasenapp, M., & Drogell, K. (2012). Elevated blood urea nitrogen/creatinine ratio is associated with poor outcome in patients with ischemic stroke. *Clinical Neurology and Neurosurgery*, 114(7), 881–884.
<https://doi.org/10.1016/j.clineuro.2012.01.031>
16. You, S., Zheng, D., Zhong, C., Wang, X., Tang, W., Sheng, L., Zheng, C., Cao, Y., & Liu, C.-F. (2018). Prognostic Significance of Blood Urea Nitrogen in Acute Ischemic Stroke. *Circulation Journal: Official Journal of the Japanese Circulation Society*, 82(2), 572–578. <https://doi.org/10.1253/circj.CJ-17-0485>
17. Lab Tests Guide. (2023). *Blood Urea VS BUN Definitions, Differences, and Similarities*. Lab Tests Guide. <https://www.labtestsguide.com/blood-urea-vs-bun>
18. Tsagalis, G., Akrivos, T., Alevizaki, M., Manios, E., Stamatellopoulos, K., Laggouranis, A., & Vemmos, K. N. (2008). Renal dysfunction in acute stroke: an independent predictor of long-term all combined vascular events and overall mortality. *Nephrology Dialysis Transplantation*, 24(1), 194–200. <https://doi.org/10.1093/ndt/gfn471>
19. Desai, A., Oh, D., Rao, E. M., Sahoo, S., Mahajan, U. V., Labak, C. M., Rohit Mauria, Shah, V. S., Nguyen, Q., Herring, E., Elder, T., Stout, A., & Shammassian, B. H. (2023). Impact of anemia on acute ischemic stroke outcomes: A systematic review of the literature. *Impact of Anemia on Acute Ischemic Stroke Outcomes: A Systematic Review of the Literature*, 18(1), e0280025–e0280025. <https://doi.org/10.1371/journal.pone.0280025>
20. Gill, D., Monori, G., Tzoulaki, I., & Dehghan, A. (2018). Iron Status and Risk of Stroke. *Stroke*, 49(12), 2815–2821. <https://doi.org/10.1161/strokeaha.118.022701>

21. Сосуинфо. *Протромбин, протромбиновый индекс и время (ПТИ, ПТВ) - понятие, роль, нормы и отклонения.* (2024). Сосудинфо.
https://sosudinfo.ru/krov/protrombin/#h2_5%20
22. Morotti, A., & Goldstein, J. N. (2020). Anticoagulant-associated intracerebral hemorrhage. *Brain Hemorrhages*, 1(1), 89–94. <https://doi.org/10.1016/j.hest.2020.01.001>
23. Sonne, J., & Lopez-Ojeda, W. (2023). *Dopamine*. Nih.gov; StatPearls Publishing.
<https://www.ncbi.nlm.nih.gov/books/NBK535451/>
24. Hoh, B. L., Ko, N. U., Sepideh Amin-Hanjani, Sherry Hsiang-Yi Chou, Cruz-Flores, S., Dangayach, N. S., Derdeyn, C. P., Du, R., Hänggi, D., Hetts, S. W., Ifejika, N. L., Johnson, R., Keigher, K. M., Leslie-Mazwi, T. M., Lucke-Wold, B., Rabinstein, A. A., Robicsek, S. A., Stapleton, C. J., Suarez, J. I., & Tjoumakaris, S. I. (2023). 2023 Guideline for the Management of Patients With Aneurysmal Subarachnoid Hemorrhage: A Guideline From the American Heart Association/American Stroke Association. *Stroke*, 54(7). <https://doi.org/10.1161/str.0000000000000436>
25. Wheble, P. C. R., Sena, E. S., & Macleod, M. R. (2007). A Systematic Review and Meta-Analysis of the Efficacy of Piracetam and Piracetam-Like Compounds in Experimental Stroke. *Cerebrovascular Diseases*, 25(1-2), 5–11. <https://doi.org/10.1159/000111493>