https://doi.org/10.15360/1813-9779-2024-3-22-31

RICD: Russian Intensive Care Dataset

Andrey V. Grechko, Mikhail Y. Yadgarov*, Alexey A. Yakovlev, Levan B. Berikashvili, Artem N. Kuzovlev, Petr A. Polyakov, Ivan V. Kuznetsov, Valery V. Likhvantsev

Federal Research and Clinical Center of Intensive Care Medicine and Rehabilitology,
25 Petrovka Str., Bldg. 2, 107031 Moscow, Russia

For citation: *Andrey V. Grechko, Mikhail Y. Yadgarov, Alexey A. Yakovlev, Levan B. Berikashvili, Artem N. Kuzovlev, Petr A. Polyakov, Ivan V. Kuznetsov, Valery V. Likhvantsev.* RICD: Russian Intensive Care Dataset. *Obshchaya Reanimatologiya = General Reanimatology.* 2024; 20 (3): 22–31. https://doi.org/10.15360/1813-9779-2024-3-22-31 [In Russ. and Engl.]

***Correspondence to:** Mikhail Y. Yadgarov, myadgarov@fnkcrr.ru

Summary

In the era of healthcare digital transformation, the scientific community faces the need for structured and available datasets for research and technological projects in the field of artificial intelligence, related to the development of new diagnostic and treatment methods.

Objective: to develop a dataset containing anonymized medical data of all patients treated at the Federal Research and Clinical Center of Intensive Care Medicine and Rehabilitology (FRCCR), and provide access for doctors and scientists of FRCCR and other centers to structured patient data for subsequent analysis and research.

Materials and Methods. The FRCCR medical information system and the tools «Asclepius», PL/SQL, Microsoft Office Excel, Power Query M, Microsoft PowerBI, Open data editor, and Python were used for data collection and representation. To provide open access to the dataset and protect the personal data of patients, the information was anonymized.

Results. We introduce the RICD (Russian Intensive Care Dataset, https://fnkcrr-database.ru/) — the first dataset of intensive care patients in the Russian Federation, developed at FRCCR based on advanced principles and methods used in international open database projects — «eICU Program» from Philips Healthcare, «MIMIC-IV», and «MIMIC-III». The developed dataset contains information on 7,730 hospitalizations of 5,115 patients (including readmissions), covering data from 3,291 hospitalizations in the intensive care units (ICUs). The total number of records in the RICD exceeds 14 million. The RICD presents medical-anthropometric data, patient movement within the institution, diagnoses, information on therapy provided, results of laboratory tests, scale assessments, and outcomes of hospitalization. RICD also contains data on several vital parameters collected from bedside monitors and other equipment of ICUs, with up to 10 evaluations per hour.

Conclusion. The RICD allows for in-depth analysis and research of clinical practices in intensive care, enabling the development of clinical decision support tools and the application of machine learning methods to enhance diagnostic tools and improve patient outcomes. With its accessibility and detailed data structure, the dataset serves as a valuable tool for both scientific research and practical applications in intensive care.

Keywords: dataset; critically ill patients; intensive care; artificial intelligence; machine learning; clinical decision support systems; https://fnkcrr-database.ru/

Conflict of interest. The authors declare a potential conflict of interest in case of commercialization of the dataset presented in this article.

Introduction

In contemporary medicine, there is a rapid development of electronic healthcare systems, which allow data collected during routine clinical practice to be organized and stored in medical institutions globally. These systems are particularly emphasized in the field of critical care due to the necessity for continuous monitoring of patients' vital functions in intensive care units (ICUs). This approach results in the generation of substantial data, which, upon detailed examination, facilitate the improvement of clinical practice. Projects such as «MIMIC-III» [1] (a database representing data from over 40,000 patients admitted to the ICU of the Beth Israel Deaconess Medical Center, USA, between 2001 and 2012) and «MIMIC-IV» [2] (data on hospitalizations of 69,653 patients from the same center, covering 2008–2019), as well as «eICU» [3] (a multicentric database comprising de-identified medical data from over 200,000 ICU admissions across numerous US medical centers between 2014 and 2015), illustrate how valuable information collected in ICUs can be utilized for scientific research and the development of new diagnostic and treatment methods. As of 2024, over 1,500 scientific publications have utilized data from the «MIMIC-III», «MIMIC-IV», and «eICU» projects, including journals such as Critical Care [4], Nature [5], Lancet [6], JAMA [7], BMC Anesthesiology [8].

Other international projects contributing significantly to clinical research include «HiRID» [9] (36,098 ICU hospitalizations in Switzerland), «AmsterdamUMCdb» [10] (23,106 ICU hospitalizations

in the Netherlands), and the database «The Children's Hospital at Zhejiang University School of Medicine» [11] (13,941 pediatric ICU admissions in China). These repositories provide unique data on patient care and clinical outcomes. However, it should be noted that due to significant differences in healthcare systems across countries, the use of foreign open database projects in domestic practice is limited. In 2023, the Siberian State Medical University introduced Russia's first clinical data repository, «SibMED Data Clinical Repository», encompassing anonymized data from more than 20,000 hospitalizations in 10 multi-specialty clinics of SibSMU [12]. As of 2024, this repository does not include monitored data of ICU patients, which limits the depth of research in intensive care.

Currently, the integration of artificial intelligence (AI) technologies into medical practice opens new possibilities for diagnosis, treatment, and monitoring of patient conditions, especially relevant for patients in intensive care settings. AI technologies encompass a broad range of methods and approaches, including neural networks, machine learning techniques, and expert systems. These tools are uniquely capable of analyzing complex and extensive datasets, including medical images, textual documents (such as diary entries, discharge summaries, clinical notes, and reports), and databases of medical parameters [13]. One of the most promising applications of AI in critical care is the real-time monitoring and analysis of patient conditions. AI technologies can predict the development of complications and adverse outcomes, identifying life-threatening conditions requiring medical staff attention, thereby becoming crucial elements of clinical decision support systems, helping to reduce the number of medical errors and improve the quality of medical care provided [14, 15]. The development of such systems requires the creation of large, structured, and accessible datasets.

The Federal Research and Clinical Center of Intensive Care Medicine and Rehabilitology (FRCCR) has implemented a concept of «digital» ICU, which entails the collection and storage on organizational servers of all data from ICU patients, including continuously monitored parameters, making it accessible for analysis. The project's goal is to develop a dataset containing anonymized medical data of all patients treated in the FRCCR, providing physicians and researchers from the FRCCR and other centers access to structured patient data for further analysis and research.

Materials and Methods

Data acquisition. The data were obtained using standard tools of the medical information system (MIS) «Asclepius», based on an Oracle 11g relational database management system. The primary dataset was extracted from various MIS modules (registration, clinical, laboratory, pharmacy, etc.) using the «Query Builder» virtual module. Additional queries were performed in the PL/SQL language directly from the relational database management system. The data extraction period was from December 2017 to July 2023. Laboratory investigation results were partly acquired by generating reports from the «Alisa» laboratory information system. Data structuring was facilitated by creating macros in the «VBA» language in Microsoft Office Excel 2021 and utilizing the «Power Query M» programming language. For demonstrating the analytical capabilities of the dataset, the BI system «Microsoft PowerBI,» the programming language «Python» in the «PyCharm 2023.2.4» environment, and IBM SPSS Statistics 27.0 were used. For standardization, directories were created to unify and standardize data from various sources. The final dataset files were created using the «Frictionless data» technology, implemented in the «Open data editor» program.

Anonymization and pseudonymization. All data were depersonalized (anonymized) to facilitate open access. The data anonymization methodology was developed based on the National Standard of the Russian Federation «Health informatics. Pseudonymization» GOST R 55036-2012/ISO/TS 25237:2008 (approved by the order of the Federal Agency on Technical Regulating and Metrology dated October 29, 2012, No. 585-st) and international open database projects. The anonymization approach was approved by the local Ethics Committee (No. 4/23/2 dated 20.12.2023). Direct and indirect patient identifiers were removed. To maintain the informativeness of the dataset, new parameters were created in each table, reflecting the timing of the investigation/examination or parameter assessment (post_admission_days and post_admission_hours). Additionally, a pseudonymization procedure was

Table 1. Descriptive characteristics of hospital and ICU patients in the RICD.

Parameter	Hospital	ICU
Number of hospitalizations	4339	3291
Number of patients*	3033	2562
Age, mean *(SD)*	59.2 (15.2)	57.8 (17.7)
Sex, male (%)	2385 (55.0)	1864 (56.6)
Length of stay, days *(Q1–Q3)*	14.0 (14.0–18.0)	32.0 (22.0–50.0)

Note. * — the number of patients exceeds 5115 because some of the patients were hospitalized in different periods of time in both the intensive care unit and the hospital.

Table 2. Brief description of the RICD tables.

Table	Number of entries	Description
1. all_patients	7730	Medical and anthropometric characteristics of patients, patient movement within the institution, hospitalization outcomes
2. ICD10_diagnoses	1058929	Diagnoses of patients according to ICD-10 classification
3. therapy_prescriptions	550880	Therapeutic assignments
4. clinical_notes	304680	Routine patient assessment and ICD-10 diagnoses
5. monitoring_data	12198188	Monitored parameters (vital signs and fluid balance parameters)
6. all_scales	19297	Scale scores
7. detailed_sofa	14859	Detailed structure of the SOFA scale score
8. complete_blood_count	43584	Complete blood count test results
9. urinalysis	37307	Results of urinalysis
10. blood_biochemistry	65882	Results of blood biochemical analysis
11. urine_biochemistry	1389	Results of biochemical analysis of urine
12. coagulation_profile	33727	Results of hemostasis system evaluation
13. acid_base_balance	6185	Results of evaluation of acid-base balance and blood gases
14. antibiotic_resistance	6794	Results of assessment of antibiotic resistance
15. bacteria_culture_test	6101	Results of bacteria culture tests
16. cerebrospinal_fluid_analysis	1021	Cerebrospinal fluid analysis

conducted, replacing identifying data (patient ID, hospitalization ID/medical history number) with pseudonyms (unique identifiers: new_patient_id, new_hosp_id), which cannot be linked to the original data without additional information stored separately.

Dataset Description

Descriptive characteristics. The RICD contains data on 7,730 hospitalizations of 5,115 patients admitted to the FRCCR during the 2017–2023 period, including 3,291 hospitalizations in the ICU (Table 1). The average age of ICU patients was 57.8 years (SD 17.7, range from 19 to 97 years), and the median duration of ICU stay was 32 days (22–50), with a maximum of 320 days. During the entire period, 405 deaths were reported. The total number of entries in all RICD tables is 14,356,553.

Data format. The dataset has been presented in several formats:

1. .csv files, interconnected through key fields (frictionless data);
2. .db format (for SQLite);
3. .pbix format (for Microsoft PowerBI system).

Additionally, metadata were provided in .json format. RICD files can be uploaded into any relational database or BI-system.

Dataset structure. The dataset comprises 16 interconnected tables (detailed description is provided in the Supplement). The unifying identifiers of all tables are: new_patient_id (modified patient ID) and new_hosp_id (modified medical history ID).

The tables contain patients' medical-anthropometric data, information about their movement within the institution, diagnoses, data on the therapy provided, results of laboratory tests, scale assessments, vital and fluid parameters assessed dynamically, and hospitalization outcomes (Table 2).

In the RICD, 85% of all entries represent assessments of monitored vital signs of ICU patients (results from pulse oximetry (SpO_2), body temperature, respiratory rate, heart rate, systolic and diastolic BP, average BP, central venous pressure) and fluid balance parameters.

Key features of RICD. 1. Availability of monitored data of ICU patients. RICD is the first dataset in the Russian Federation that presents data collected from bedside monitors and other equipment of the ICU. The sampling frequency (data assessments) is up to 10 assessments per hour.

2. The primary cohort of patients consists of individuals who have been staying in ICUs for extended periods. More than 60% of the patients had a reduced level of consciousness (Glasgow Coma Scale score <15 points).

3. Availability of bacteria culture tests with evaluation of antibiotic resistance.

Technical validation. For the assessment of the developed dataset, an interdisciplinary team was involved, which conducted data validation, evaluated the integrity of data through cross-checking tables, identified acceptable ranges of data values, and determined the completeness of data anonymization.

Access platform. To provide access to the RICD project, a platform (website: https://fnkcrr-database.ru) was developed. The website presents complete information about the RICD project, describes the structure of the dataset, and offers the possibility to obtain a demo version of the dataset, representing the hospitalization data of 10 patients (over 60,000 monitored parameter assessments in the ICU). Additionally, materials on working with RICD biomedical data are available, which can be used to study the dataset and familiarize with possible tools for its analysis.

Applying for access offers the option «participate in the RICD project», which creates a basis for the integration and unification of dataset of various medical centers.

Demonstration of analytical capabilities. Figure 1 presents the parameters of a patient monitored over 15 days of hospitalization at our center. During

the specified period, there are over 1000 assessments for each of the vital parameters (including SpO_2, respiratory rate, and heart rate), dynamics of changes in laboratory parameters, and evaluations based on clinical scales. Moreover, the RICD provides the ability to evaluate the mechanical ventilation status and therapy prescriptions, including the use of vasopressor and inotropic drugs.

Conclusion

The presented RICD dataset enables in-depth analysis and research of clinical practices in intensive care, development of clinical decision support tools and application of machine learning methods to solve diagnostic problems and improve patient outcomes. Due to the accessibility and detailed structuring of the data, the dataset will be a useful tool for both scientific and practical applications. The RICD project is recommended for use by researchers, data science and machine learning specialists, and developers of advanced digital health solutions.

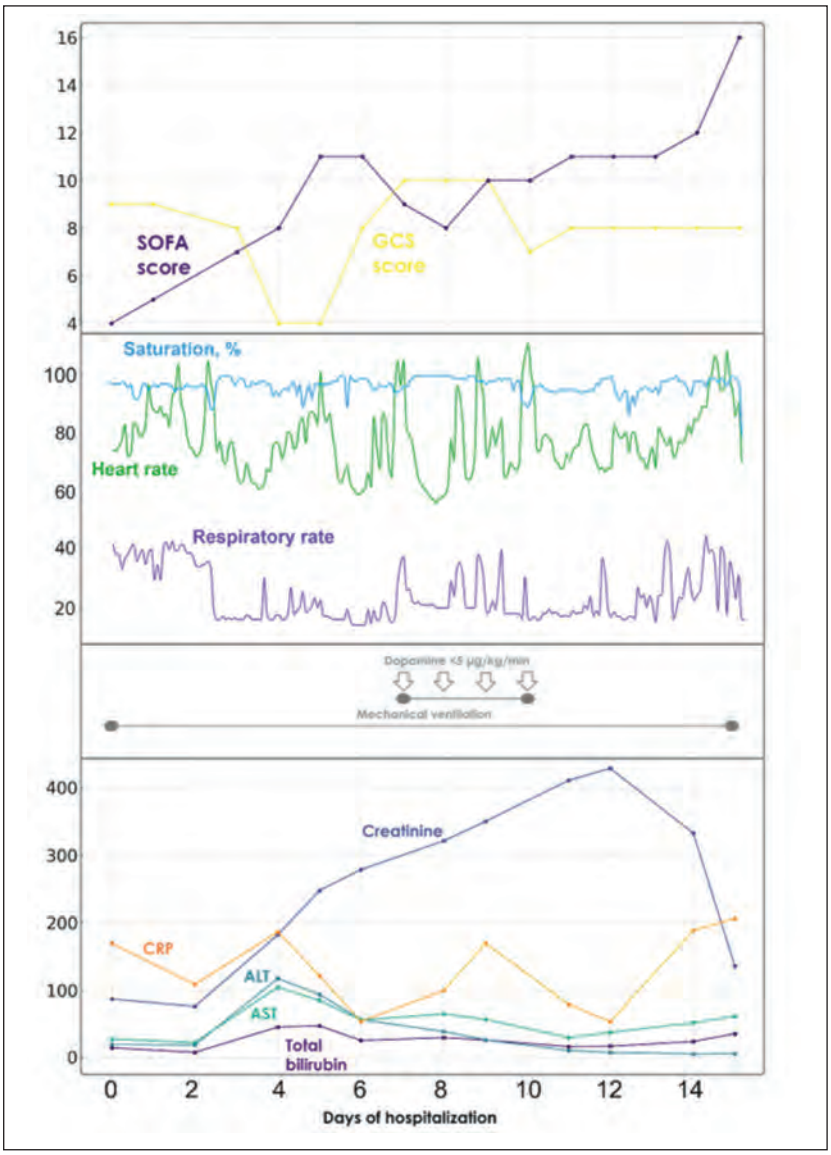

Fig. Parameters monitored during 15 days of the patient's hospitalization at Federal Research and Clinical Center of Intensive Care Medicine and Rehabilitology.
Note. Data from tables all_scales, monitoring_data, blood_biochemistry, and detailed_SOFA scores were used.

Supplement. Detailed description of the RICD tables

1. All_patients (7,730 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
sex	Patient gender	—
body	Weight at admission	kg
height	Patient's height at admission	cm
BMI	Body mass index at admission	kg/m²
age	Patient's age at admission	years
patient_condition	Patient's condition at admission	—
transfer	Fact of transfer from another institution	—
adm_year	Year of admission	—
admission_department	Admission department	—
discharge department	Discharge department	—
ICU_patients	ICU patient	—
length_of_stay	Length of stay	days
fatal_outcome	Fatal outcome	—

2. ICD10_diagnoses (1,058,929 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
document_type	Document type	—
diagnosis_type	Diagnosis type	—
ICD_10	ICD-10 code	—

3. therapy_prescriptions (550,880 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
prescription_rus	Prescription in Russian	—
prescription_eng	Prescription in English	—
Time (12:00_am — 11:00_pm)	Estimated time of performance/assignment status	V assigned + completed X canceled – failed

4. clinical_notes (304,680 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
patient_condition	Patient's condition at the time of assessment	—
ICD_10	ICD-10 codes	—

5. monitoring_data (12,198,188 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
parameter	Monitored parameter*	—
unit	Units	—
hour	Evaluation time	—
value	Value	—

***Vital parameters:** saturation (SpO_2), temperature, respiratory rate, heart rate, diastolic BP, systolic BP, mean AP, body mass, glucose, respiratory volume, central venous pressure, BIS, $EtCO_2$. ***Fluid parameters:** fluid intake per os, diuresis, enteral feeding, infusion, liquor, stool/stoma, other intake, other output, gastrostoma, drainages volume, nasogastric tube/vomitus, cystostomy.

6. all_scales (19,297 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
scales	Scales* and evaluation results	—

***Scales:** CRS-R, APACHE II, CHA_2DS_2-VASc, DRS, FIM, FOUR, GRACE, HAS-BLED, NIHSS, POSSUM, SAPS II, SOFA, Barthel, Waterlow (pressure ulcer risk), Geneva score, Glasgow coma scale, Classification of surgical and anesthesia risk (MNOAR), Ashworth scale, Caprini (DVT/PE risk), Palliative performance scale, Rehabilitation routing scale, Rivermead mobility index, Modified Rankin scale, Wells' Criteria for DVT.

7. detailed_sofa (14,859 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
report	Text Conclusion	—
sofa_score	SOFA score	—
FiO_2 %	FiO_2 %	%
PaO_2	PaO_2	mm Hg
PaO_2/FiO_2	PaO_2/FiO_2	—
mechanical_ventilation	Mechanical ventilation	—
platelets	Platelets	10^9/L
GSC_score	Glasgow Coma Scale Score	—
bilirubin	Bilirubin	umol/L
systolic_AP	Systolic blood pressure	mm. hg
diastolic_AP	Diastolic blood pressure	mm. hg
mean_AP	Mean blood pressure	mm. hg
vasoactive_drugs	Use of vasoactive drugs	—
creatinine	Creatinine	umol/L
daily_diuresis	Daily urine output	ml

8. complete_blood_count (43,584 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
WBC	White Blood Cells	10^9/L
RBC	Red Blood Cells	10^{12}/L
HGB	Hemoglobin	g/L
HCT	Hematocrit	%
MCV	Average red blood cell volume	fl
MCH	Average red blood cell content	pg
MCHC	Mean red blood cell hemoglobin concentration	g/l
RDW	Red blood cell distribution width by volume	%
RDW-SD	Standard deviation of red blood cell distribution width by volume	fl
PLT	Platelets	10^9/L
PCT	Thrombocrit	%
MPV	Average platelet volume	fl
PDW	Platelet distribution width by volume	%
NEU%	Neutrophil percentage	%
NEU	Absolute neutrophil count	10^9/L
LYM %	Percentage of lymphocytes	%
LYM	Absolute number of lymphocytes	10^9 /L
MONO %	Percentage of monocytes	%
MONO	Absolute number of monocytes	10^9/L
EOS %	Percentage of eosinophils	%
EOS	Absolute number of eosinophils	10^9/L
BASO %	Percentage Basophil percentage	%
BASO	Absolute basophil count	10^9/L
NRBC%	Percentage of normoblasts	%
NRBC #	Absolute number of normoblasts	10^9/L
%RETIC	Percentage of reticulocytes	%
RETIC	Absolute number of reticulocytes	10^9/L
IRF	Immature reticulocyte fraction	%
MRV	Average reticulocyte volume	fl
band_neutrophil	Band neutrophils	%
segmented_neutrophil	Segmented neutrophils	%
ESR	Erythrocyte sedimentation rate	mm/h

9. urinalysis (37,307 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
urine_color	Urine color	—
clarity/turbidity	Urine transparency	—
specific_gravity	Specific gravity	units
PH	pH of urine	—
protein	Protein	mmol/L
glucose	Glucose	mmol/L
nitrites	Nitrites	—
ketones	Ketones	mmol/L
bilirubin	Bilirubin	umol/L
ascorbic_acid	Ascorbic acid	mmol/L
urobilinogen	Urobilinogen	umol/L
squamous_epithelial_cells	Squamous epithelium	—
transitional_epithelial_cells	Transitional epithelium	—
WBCs	WBCs	—
RBCs	RBCs	—
bacteria	Bacteria	—
hyaline_casts	Hyaline casts	—
nonclassified_casts	Nonclassified casts	—
mucus	Mucus	—

10. blood_biochemistry (65,882 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique Hospitalization ID	—
post_admission_days	Days Since Admission	—
post_admission_hours	Hours Since Admission	—
HbA1C_NGSP	Glycated Hemoglobin Method NGSP	%
troponin	Troponin (I, T)	pg/mL
total_bilirubin	Total Bilirubin	µmol/L
direct_bilirubin	Direct Bilirubin	µmol/l
total_protein	Total Protein	g/l
albumin	Albumin	g/l
prealbumin	Prealbumin	g/l
urea	Urea	mmol/L
creatinine	Creatinine	µmol/L
glucose	Glucose	mmol/l
triglycerides	Triglycerides	mmol/l
cholesterol	Cholesterol	mmol/l
HDL	High-density Lipoprotein	mmol/l
LDL	Low-density Lipoprotein	mmol/l
atherogenic_coefficient	Atherogenic Coefficient	—
VLDL	Very Low-density Lipoprotein	mmol/l
magnesium	Magnesium	mmol/l
calcium	Calcium	mmol/l
phosphorus	Phosphorus	mmol/l
iron	Iron	µmol/l
latent_iron_binding_capacity	Latent Iron-binding Capacity of Serum	µmol/l
transferrin	Transferrin	mg/dL
potassium	Potassium	mmol/l
sodium	Sodium	mmol/L
procalcitonin	Procalcitonin	ng/mL
chlorides	Chlorides	mmol/L
LDH	Lactate Dehydrogenase	U/L
ALT	Alanine Aminotransferase	U/L
AST	Aspartate Aminotransferase	U/L
GGT	Gamma-glutamyl Transferase	U/L
alkaline_phosphatase	Alkaline Phosphatase	E/l
amylase	Amylase	Unit/l
CRP	C-reactive Protein	mg/l
uric_acid	Uric acid	µmol/L
rheumatoid_factor	Rheumatoid Factor	U/mL
cholinesterase	Cholinesterase	U/L
creatine_kinase	Creatine Kinase	U/L
anti_streptolysin_O	Antistreptolysin O	U/mL
HBDH	Hydroxybutyrate Dehydrogenase	U/L
CK_MB	Creatine Kinase MB	U/l
APO_A1	Apolipoprotein A1	g/l
APO_B	Apolipoprotein In	g/l
APO_B_APO_A1_ratio	The Ratio of ApoB/ApoA1	—
ferritin	Ferritin	µg/l

11. urine_biochemistry (1,389 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
Ca+24	Calcium in daily urine	mmol/day
Cl+24	Chlorides in daily urine	mmol/day
CREAT24	Creatinine in the daily urine	mmol/day
GLU24	Glucose in the daily urine	g/day
K+24	Potassium in daily urine	mmol/day
Mg24	Magnesium in daily urine	mmol/day
Na+24	Sodium in daily urine	mmol/day
urine_albumin_24h	Albumin in the daily urine	µg/day
24h_urine_urea	Urea in the daily urine	mmol/day
urea_in_urine	Urea in the urine	mmol/day
nitrogen_loss_per_day	Nitrogen loss per day	g/day
phosphorus24	Phosphorus in daily urine	mmol/day
urine_volume	Urine Volume in 24 hours	l/day
urine_glucose	Glucose in the urine	mmol/l
microalbumin	Microalbumin in the urine	mg/l
microalbumin24	Microalbumin in daily urine	mmol/day
total_protein_urine	Total protein in urine	g/day
PROT24	Protein in the daily urine	mg/day
urine_calcium	Calcium in the urine	mmol/L
urine_magnesium	Magnesium in the urine	mmol/L
uric_acid_urine	Uric acid in the urine	mg/l
urine_sodium	Sodium in the urine	mmol/l
urine_phosphorus	Phosphorus in the urine	mmol/l
urine_chlorides	Chlorides in the urine	mmol/l
urine_potassium	Potassium in the urine	mmol/l
urine_amylase	Amylase in the urine	U/L
urine_creatinine	Creatinine in the urine	mg/l

12. coagulation_profile (33,727 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
anti-Xa	Anti-Xa	U/ml
PT_seconds	Prothrombin Time	sec
quick_test _prothrombin_time	Quick test prothrombin time	%
prothrombin_ratio	Prothrombin ratio	—
international_normalized_ratio	International normalized ratio	—
fibrinogen_calculated	Calculated fibrinogen	g/l
fibrinogen_Clauss	Clauss fibrinogen	g/l
activated_partial_thromboplastin_time	Activated partial thromboplastin time	sec
thrombin_time	Thrombin time	sec
D-dimer	D-dimer	mg/L

13. acid_base_balance (6,185 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
A-aDO_2	Alveolar-arterial oxygen gradient	mm Hg
temperature_corrected_pH	Temperature-corrected venous pH	—
temperature_corrected_pO_2	Venous temperature-corrected pO_2	mm Hg
respiratory_index	Respiratory index	—
pH	Venous pH	—
pH_arterial	Arterial pH	—
pCO_2	pCO_2 venous	mm Hg
pCO_2_arterial	Arterial pCO_2	mm Hg
pCO_2_capillary	Capillary pCO_2	mm Hg
temperature_corrected_pCO_2	Temperature-corrected pCO_2 venous	mm Hg
pO_2	pO_2 venous	mm Hg
pO_2_arterial	Arterial pO_2	mm Hg
Na+	Sodium	mmol/L
K+	Potassium	mmol/L
Ca++	Calcium	mmol/L
Glu	Glucose	mmol/L
Hct	Hematocrit	%
Lac	Lactate	mmol/L
total_hemoglobin	Total hemoglobin	g/L
sO_2_arterial	Arterial oxygen saturation	%
BE(B)	Base excess (venous blood)	mm Hg
Beecf	Base excess (extracellular fluid)	mmol/L
paO_2/pAO_2	paO_2/pAO_2	—
$\%FiO_2$	$\%FiO_2$	%
HCO_3^-_std	Standard HCO_3^-	—
HCO_3^-	HCO_3^-	mmol/L
P/F Ratio	P/F Ratio	—
temp	Temperature	C

14. antibiotic_resistance (6,794 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
microorganism	Microorganism (pathogen)	—
biological_material	Biological material	—
antibiotics (MIC)*	Estimation of the minimum inhibitory concentration for each antibiotic	—
antibiotics (RSI)*	Evaluation of antibiotic resistance shown as R, S and I	R (Resistant) S (Sensitive) I (Intermediate

***Antibiotics:** Aztreonam, Amikacin, Amoxicillin/Clavulanate (f), Ampicillin, Gentamicin, Imipenem, Colistin, Meropenem, Nitrofurantoin, Norfloxacin, Piperacillin/Tazobactam, Trimethoprim/Sulfamethoxazole, Fosfomycin with Glucose-6-Phosphate, Cefepime, Cefixime, Ceftazidime, Ceftriaxone, Ciprofloxacin, Ertapenem, Amoxicillin/Clavulanate, Vancomycin, Gentamicin-syn, Daptomycin, Clindamycin, Levofloxacin, Linezolid, Oxacillin, Penicillin G, Rifampin, Streptomycin-synergism, Teicoplanin, Tetracycline, Tigecycline, Tobramycin, Fusidic Acid, Quinupristin/Dalfopristin, Cefoxitin, Erythromycin, Netilmicin, Piperacillin, Cefuroxime, Cefoperazone, Cefotaxime, Amoxicillin, Moxifloxacin, Pristinamycin, Chloramphenicol, BMS-284756, Cefazolin, High-activity Mupirocin, Doxycycline, Mupirocin, Trimethoprim, Clarithromycin, Cefoperazone/Sulbactam, Ofloxacin, Ampicillin/Sulbactam (f), Ceftolozane-Tazobactam, Polymyxin B, Ceftazidime-Avibactam, Kanamycin, Kanamycin-syn, Moxalactam, Ceftaroline, Cefepime/Sulbactam.

15. bacteria_culture_test (6,101 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
biomaterial_rus	Biological material in Russian	—
biomaterial_eng	Biological material in English	—
microorganism*	Microorganism (pathogen)	CFU/ml

*__Microorganisms:__ *A. baum/haem, Chry. indolog, Chrys. meningosept, Pseud. oryzihabit, Ser. liquefaciens, Cory. striatum, Ped. pentosaceus, Staph. schleiferi, Strep. dysgal./ca, Strep. gallolytic, Strep. mitis gr., Strep. parasangui, Strep. vestibular, G. haemolysan, M. lacunata, Moraxella. sp., A. baum/calc. comp, A. baumannii, A. lwoffi/haemolyt, Acinetobac. sp., Coryn. amycolatum, Coryn. urealytic, Corynebac. sp., Corynebacterium, Bacil. cereus, P. putida, R. radiobacte, Alcalig. faecalis, S. aureus, S. epidermidi, S. haemolytic, S. capitis, S. coh-ss-coh, S. xylosus, Staphyl. carnosus, S. agalactiae, S. constellatus, Str. anginosus, Str. gordonii, Str. intermedius, Str. mitis, Str. oralis, Str. pneumoniae, Str. sanguinis, Leuconost. pseudom, Achromobacter ssp., B. cepacia CF, Bur. cepacia, Bur. gladioli, S. maltophili, Sph. paucimob, K. oxytoca, K. pne-ss-oza, K. pne-ss-pne, K. pneu/oxy, Kleb. pneumoniae, E. coli, E. coli urea+, E. cloacae, E. gergoviae, P. mirabilis, P. penn/vulg, P. penneri, P. vulgaris, S. marcescens, S. odorifera1, S. plymuthica, A. caviae, Lact. catenafor, H. alvei, P. multocida, C. violaceum, C. davisae, C. lapagei, C. neteri, C. freundii, C. koseri, K. ascorbata, Morganella morgan, P. agglomeran, P. alcalafaci, P. rettgeri, P. rustigian, P. stuartii, E. avium, E. casselifla/gall, E. faecalis, E. faecium, E. raffinosus, M. wisconsens, Klebsiella pneumo, Streptococcus con, Candida albicans, E. aerogenes, Pseudomonas aeruginosa.*

16. cerebrospinal_fluid_analysis (1,021 records)

Fields	Description	Optional
new_patient_id	Unique Patient ID	—
new_hosp_id	Unique hospitalization ID	—
post_admission_days	Days since admission	—
post_admission_hours	Hours since admission	—
lymphocytes_csf	Lymphocytes in the cerebrospinal fluid	N/in the field of view
eosinophils_csf	Eosinophils in the cerebrospinal fluid	N/in the field of view
neutrophils_csf	Neutrophils in the cerebrospinal fluid	N/in the field of view
macrophages_csf	Macrophages in the cerebrospinal fluid	N/in the field of view
monocytes_csf	Monocytes in the cerebrospinal fluid	N/in the field of view
arachnoid cells_csf	Arachnoid cells in the cerebrospinal fluid	N/in the field of view
granular spheres_csf	Granular spheres in the cerebrospinal fluid	N/in the field of view
erythrocytes_csf	Erythrocytes in the cerebrospinal fluid	N/in the field of view

References

1. *Johnson A.E.W., Pollard T.J., Shen L., Lehman L.W.H., Feng M., Ghassemi M., Moody B., et al.* MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016; 3: 160035. DOI: 10.1038/sdata.2016.35. PMID: 27219127.
2. *Johnson A.E.W., Bulgarelli L., Shen L., Gayles A., Shammout A., Horng S., Pollard T.J., et al.* MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023; 10 (1): 1. DOI: 10.1038/s41597-022-01899-x. PMID: 36596836.
3. *Pollard T.J., Johnson A.E.W., Raffa J.D., Celi L.A., Mark R.G., Badawi O.* The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data*. 2018; 5: 180178. DOI: 10.1038/sdata.2018. 178. PMID: 30204154.
4. *Chen H., Zhu Z., Zhao C., Guo Y., Chen D., Wei Y., Jin J.* Central venous pressure measurement is associated with improved outcomes in septic patients: an analysis of the MIMIC-III database. *Crit Care*. 2020; 24 (1): 433. DOI: 10.1186/S13054-020-03109-9. PMID: 32665010.
5. *Arévalo A.R., Maley J.H., Baker L., da Silva Vieira S.M., da Costa Sousa J.M., Finkelstein S., Mateo-Collado R., et al.* Data-driven curation process for describing the blood glucose management in the intensive care unit. *Sci Data*. 2021; 8 (1): 80. DOI: 10.1038/s41597-021-00864-4. PMID: 33692359.
6. *Liu P., Li S., Zheng T., Wu J., Fan Y., Liu X., Gong W., et al.* Subphenotyping heterogeneous patients with chronic critical illness to guide individualised fluid balance treatment using machine learning: a retrospective cohort study. *EClinicalMedicine*. 2023; 59: 101970. DOI: 10.1016/j.eclinm.2023. 101970. PMID: 37131542.
7. *Tyler P.D., Du H., Feng M., Bai R., Xu Z., Horowitz G.L., Stone D.J., et al.* Assessment of intensive care unit laboratory values that differ from reference ranges and association with patient mortality and length of stay. *JAMA Netw Open*. 2018; 1 (7): e184521. DOI: 10.1001/JAMANETWORKOPEN.2018.4521. PMID: 30646358.
8. *Sun Y., He Z., Ren J., Wu Y.* Prediction model of in-hospital mortality in intensive care unit patients with cardiac arrest: a retrospective analysis of MIMIC — IV database based on machine learning. *BMC Anesthesiol*. 2023; 23 (1): 178. DOI: 10.1186/S12871-023-02138-5. PMID: 37231340.
9. *Hyland S.L., Faltys M., Hüser M., Lyu X., Gumbsch T., Esteban C., Bock C., et al.* Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat Med*. 2020; 26 (3): 364–373. DOI: 10.1038/s41591-020-0789-4. PMID: 32152583.
10. *Thoral P.J., Peppink J.M., Driessen R.H., Sijbrands E.J.G., Kompanje E.J.O., Kaplan L., Bailey H., et al.* Sharing ICU patient data responsibly under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) example. *Crit Care Med*. 2021; 49 (6): e563–e577. DOI: 10.1097/CCM.0000000000004916. PMID: 33625129.
11. *Zeng X., Yu G., Lu Y., Tan L., Wu X., Shi S., Duan H., et al.* PIC, a paediatric-specific intensive care database. *Sci Data*. 2020; 7 (1): 14. DOI: 10.1038/s41597-020-0355-4. PMID: 31932583.
12. *Куликов Е.С., Федорова О.С., Толмачев И.В., Рязанцева У.В., Вражнов Д.А., Губанов А.В., Нестерович С.В., с соавт.* Русскоязычный репозиторий открытых клинических данных SibMED Data Clinical Repository. *Бюллетень Сибирской Медицины*. 2023; 22 (2): 182–184. *Kulikov E.S., Fedorova O.S., Tolmachev I.V., Ryazantseva U.V., Vrazhnov D.A., Gubanov A.V., Nesterovich S.V., et al.* Russian-language repository of the open clinical data «SibMed Data Clinical Repository». *Bulletin of Siberian Medicine=Bulleten Sibirskoy Meditsiny*. 2023; 22 (2): 182–184. (in Russ.). DOI: 10.20538/1682-0363-2023-2-182-184.
13. *Davenport T., Kalakota R.* The potential for artificial intelligence in healthcare. *Futur Healthc J*. 2019; 6 (2): 94–98. DOI: 10.7861/futurehosp.6-2-94. PMID: 31363513.
14. *Реброва О.Ю.* Жизненный цикл систем поддержки принятия врачебных решений как медицинских технологий. *Врач и Информационные Технологии*. 2020; (1): 27–37. *Rebrova O.Yu.* Life cycle of decision support systems as medical technologies. *Doctor and Information Technology=Vrach i Informatsionnye Tekhnologii*. 2020; (1): 27–37. (in Russ.). DOI: 10.37690/1811-0193-2020-1-27-37.
15. *Гусев А.В., Зарубина Т.В.* Поддержка принятия врачебных решений в медицинских информационных системах медицинской организации. *Врач и Информационные Технологии*. 2017; (2). УДК 614.2. *Gusev A.V., Zarubina T.V.* Clinical decision support in medical information systems of a medical organization. *Doctor and Information Technology=Vrach i Informatsionnye Tekhnologii*. 2017; (2). (in Russ.). UDC 614.2
16. *Shepherd A.* Frictionless Data 2016. https://www.frictionlessdata.io/ (accessed 3 March 2024).

Received 21.03.2024
Accepted 29.03.2024