УНИВЕРСИТЕТ ИТМО
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2022 Том 22 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2022 Vol. 22 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)
НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2022-22-1-155-166

Imputation and system modeling of acid-base state parameters for different groups of patients

Dmitry I. Kurapeev[1], Mikhail S. Lushnov[2], Tianxing Man[3], Natalia A. Zhukova[4]✉

[1,2] Almazov National Medical Research Centre, Saint Petersburg, 197341, Russian Federation
[3,4] ITMO University, Saint Petersburg, 197101, Russian Federation
[4] St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

[1] dkurapeev@gmail.com, https://orcid.org/0000-0002-2190-1495
[2] Lushnov_ms@almazovcentre.ru, https://orcid.org/0000-0002-9683-1858
[3] mantx626@gmail.com, https://orcid.org/0000-0003-2187-1641
[4] nazhukova@mail.ru✉, https://orcid.org/0000-0001-5877-4461

Abstract
The paper investigated the possibility of correct replacement of missing values in sets of acid-base state in the artery and vein in different groups of patients with different outcomes of the disease: “discharged”, “died”, “transferred to another medical institution”, as well as prospects for the application of individual optimization multidimensional estimates of these biomedical parameters in the form of projections on a one-dimensional space. The relevance of the above tasks is determined by the need for the full use of medical data in the analysis of large repositories of information of medical organizations and the provision of verified multidimensional assessments of biomedical systems to doctors from a large range of patient health indicators. A statistical method has been applied to verify the correctness of imputation data sets using discriminant analysis procedures. Further, the imputed data set was processed to obtain a symmetric correlation matrix optimized in a certain way and the accompanying logarithms of criterion functions that are individual system assessments of the condition of each patient in different groups of patients at a certain point in the study. After that, to identify differences in the logarithms of the criterion functions of the acid-base state parameters between groups of patients with different outcomes, the authors used the method of calculating multidimensional Hotelling T^2 statistics. The correctness of the application of discriminant analysis procedures to verify the imputation of data sets is shown. Differences in the logarithms of the criteria functions of the acid-base state indicators between venous and arterial blood by patient outcome groups were revealed. Significant differences in the parameters of acid-base state based on the multidimensional statistics of T^2 Hotelling between groups of patients with different outcomes were revealed. It is found that data imputation significantly increases the volume and representativeness of the sample under study. It is demonstrated that the substituted data make it possible to carry out a systematic statistical assessment of the totality of body parameters based on the calculation of the logarithms of the criterion functions of the acid-base state. Such logarithms make it possible to reliably distinguish patients in different groups of patients by outcomes in three groups: “discharged”, “deceased”, “transferred to another medical institution”. 100 % differences of biochemical parameters according to the multivariate T^2 Hotelling statistics between these three groups of patients with COVID-19 are shown. The results of the study can be applied in the development of information systems of individual medical biochemical and hematological devices and analyzers and corresponding artificial intelligence systems in the future.

Keywords
acid-base state, imputation, medical information system, COVID-19, criterion function

Acknowledgements
The research was supported by the Russian Science Foundation under Grant No. 17-15-01177.
The research was carried out within the framework of the budget theme No. 0060-2019-0011.

For citation: Kurapeev D.I., Lushnov M.S., Man T., Zhukova N.A. Imputation and system modeling of acid-base state parameters for different groups of patients. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 1, pp. 155–166. doi: 10.17586/2226-1494-2022-22-1-155-166

УДК 616.092+616.98, 303.425.6, 338.45, 519.873, 681.518

Вменение и системное моделирование параметров кислотно-основного состояния различных групп пациентов

Дмитрий Ильич Курапеев[1], Михаил Степанович Лушнов[2], Тяньсин Ман[3], Наталья Александровна Жукова[4]✉

[1,2] Национальный медицинский исследовательский центр имени В.А. Алмазова, Санкт-Петербург, 197341, Российская Федерация
[3,4] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[4] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] dkurapeev@gmail.com, https://orcid.org/0000-0002-2190-1495
[2] Lushnov_ms@almazovcentre.ru, https://orcid.org/0000-0002-9683-1858
[3] mantx626@gmail.com, https://orcid.org/0000-0003-2187-1641
[4] nazhukova@mail.ru✉, https://orcid.org/0000-0001-5877-4461

Аннотация

Предмет исследования. Исследована возможность корректной замены недостающих значений в наборах кислотно-основного состояния в артерии и вене в разных группах пациентов с различными исходами заболевания: «выписан», «умер», «переведен в другое медицинское учреждение». Рассмотрены перспективы применения индивидуальных оптимизационных многомерных оценок данных биомедицинских параметров в виде проекций на одномерное пространство. Решение поставленных задач необходимо для полноценного использования медицинских данных при анализе больших хранилищ информации медицинских организаций и предоставления врачам из большого круга показателей о здоровье пациентов верифицированных многомерных оценок биомедицинских систем. **Методы.** Применен статистический метод для проверки корректности в наборах данных вменения с использованием процедур дискриминантного анализа. Выполнена оценка импутированного набора данных для получения оптимизированной симметричной корреляционной матрицы и сопутствующих логарифмов критериальных функций. Получены индивидуальные системные оценки состояния каждого пациента в разных группах пациентов в определенный момент исследования. Применен метод вычисления многомерной статистики Хотеллинга T^2 для выявления различий в логарифмах критериальных функций параметров кислотно-основного состояния между группами пациентов с различными исходами. **Результаты.** Показана корректность применения процедур дискриминантного анализа для проверки вменения наборов данных. Выявлены статистически значимые отличия логарифмов критериальных функций показателей кислотно-основного состояния между венозной и артериальной кровью и биохимических параметров крови на основе многомерной статистики Хотеллинга T^2 между группами пациентов с различными исходами. **Практическая значимость.** Доказано, что импутация данных значительно увеличивает объем и представительность исследуемой выборки. Продемонстрировано, что замещенные данные позволяют проводить системную статистическую оценку совокупности параметров организма на основе расчета логарифмов критериальных функций кислотно-основного состояния. Такие логарифмы позволяют точно различать пациентов по исходам в трех группах: «выписанные», «умершие», «переведенные в другое медицинское учреждение». Показаны 100 % различия биохимических показателей по многомерной статистике Хотеллинга T^2 между указанными тремя группами пациентов с COVID-19. Результаты исследования могут быть применены при разработке информационных систем отдельных медицинских биохимических и гематологических приборов и анализаторов и в перспективе соответствующих систем искусственного интеллекта.

Ключевые слова

кислотно-основное состояние, вменение, медицинская информационная система, COVID-19, критериальная функция

Благодарности

Исследование выполнено при поддержке Российского научного фонда в рамках гранта № 17-15-01177.
Исследование проводилось в рамках бюджетной темы № 0060-2019-0011.

Ссылка для цитирования: Курапеев Д.И., Лушнов М.С., Ман Т., Жукова Н.А. Вменение и системное моделирование параметров кислотно-основного состояния различных групп пациентов // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 22, № 1. С. 155–166 (на англ. яз.). doi: 10.17586/2226-1494-2021-22-1-155-166

Introduction

Nowadays, there are currently millions of records accumulated in medical information systems (MIS) and databases. Data tend to contain a significant number of omissions. Medical data can be analyzed from physiological system positions [1] and synergetics [2]. In the first case, when considering MIS, the term “system” is understood primarily from the technical point of view as data accumulation and storage, and in the second case, it is used as a medical and physiological concept for the meaningful modeling of the conditions of patients with various diseases. Therefore, for the qualitative modeling of the systemic conditions of patients, it is important to take into account the maximum possible amount of data without their distortion and loss of information since the system model implies the most extensive filling of data sets. The results of the research and scientific ideas of academician

P.K. Anokhin [1] are taken as the basis for the description of physiological systems in this work.

The aim of the work is to conduct systematic studies that are based on correctly replaced missing values and the assessment of the average criterion functions (CF) of a set of parameters of the acid-base state of arterial and venous blood in different groups of patients (DGP) [3–8].

Materials and Methods

The calculation of the criterion functions (CF) is carried out based on the results of measurements for many individual parameters that characterize the state of the system under study. A correlation matrix is constructed for the entire sample of the biosystem, which is subjected to a special transformation using branches and boundaries with the selection of the optimal subset of features and the evaluation of the criterion function for each patient. The method is based on the estimation of some monotone function — CF from some biological set (**A**). The algorithm is based on calculating the maximum CF based on a certain quadratic form and on finding the largest set of n variables that maximizes the CF for the entire subset containing m features.

CF is calculated using the quadratic form: $\mathbf{C}(\mathbf{A}_m) = (\mathbf{X}_m{}^T)\mathbf{S}_m{}^{-1}(\mathbf{X}_m)$, where $\mathbf{A}_m$ is a set of m variables, $\mathbf{X}_m$ is vector of variables (a set of bioparameters — the functional system of a particular patient) and $\mathbf{S}_m$ is symmetric positive definite correlation matrix of size $m \times m$; symbol $\mathbf{X}_m{}^T$ means the operation of transposing a vector, $\mathbf{S}_m{}^{-1}$ is the operation of calculating the inverse matrix [8].

Thus, the above information shows that correctly constructed matrices without missing data can ensure successful modeling.

The relevance of the problem focusing on the influence of omissions in data on the results and conclusions of medical research is noted in a large number of publications [9, 10].

In the field of medical statistics (the generally accepted Russian and English terminology has not been defined yet) the following are the terms used to search for publications: data with omissions, omissions or omitted values (observations), censored values (observations), censorship (the process of forming censored data and data with omissions), random omissions, non-random omissions, non-random censorship (that is, a non-random mechanism for forming omissions and censored observations), offset (in the results and the conclusions of the study), the loss of part of the data, imputation (replacement of the missing value by its assessment) [11]. The omissions in the datasets can be processed using statistics, for example, using multivariate analysis.

The problem of omissions and the resulting shifts in research output is most relevant in the branches of medicine dealing with extreme and terminal conditions, that is, in traumatology [12, 13], resuscitation [14], emergency cardiology [15, 16]. The problem of data loss in experimental and clinical pharmacology studies remains relevant [17]. In the USA, a special regulatory act has been issued [18], which defines 3 gradations of omissions in clinical pharmacology data, depending on the degree of their randomness, and gives instructions for their statistical processing. However, some experts in this field express the opinion that this is not recognized as a serious problem, or consider it a nuisance that should be ignored [16].

Usually, the data of sociological research obtained through interviews are characterized by a large number of omissions [19]. To work effectively with them, it becomes necessary to fill in the gaps in the data since simply discarding observations containing missing values can lead to changes in the statistical characteristics of the sample [20].

Despite many years of development of methods for analyzing data with omissions, satisfactory solutions for many medical problems have not been found yet. Most authors recognized that there is no universal algorithm for statistical and mathematical processing of data with omissions [17, 21]. As a result, for example, in the work [22], 6 methods of data processing with omissions were used in parallel.

One of the most common methodological approaches is the imputation (replacement) of the missing value by its statistical assessment, that is, the value obtained from the preserved (non-missed) values that are closest to the missed one. Among such methods, multiple imputation is very promising, and it is used, in particular, in the works [13, 21, 23, 24].

The implementation of the MICE algorithm (R-package) made it possible to form the database necessary for the study with the restored values [25]. The obtained data sets are used to solve the problem of modeling the dependence of the labor supply on the individual's health characteristics using spatial regression with fixed effects on panel data. It is shown that filling in the gaps in data makes it possible to eliminate some obstacles that arise during econometric modeling. In addition, excluding observations with omissions would probably lead to a bias in parameter estimates [26, 27].

Thus, from the conducted literature review, it can be concluded that in recent years, methods for statistical and mathematical analysis of data with omissions have been intensively developed.

Ethical considerations. In this paper, studies of depersonalized acid-base state (ABS) data of DGP were conducted. The purpose of the study was not to make contacts with patients and to store and process their personal data.

Results

When analyzing the biochemical parameters in DGP (patients with COVID-19), a systematic approach to functional systems was applied according to [2]. At the beginning of March 2021, 1,687 such patients were registered in the MIS qMS of the Almazov National Medical Research Center (Saint Petersburg, Russia). Among them, 3 groups of patients were identified: "discharged", "died", and "transferred to another medical facility (MF)".

The gender and age composition of the patients was as follows. There were about the same number of men and women. Men 21–35 years — 6 %, 35–60 years —

42 %, 60–75 years — 35 %, 75–90 years — 16 %. Women 20–35 years — 6 %, 35–55 years — 28 %, 55–75 years — 49 %, 75–90 years — 16 %.

ABS samples were made to patients as needed according to the indications and prescriptions of the medical staff. From the database, 18,658 studies of ABS of patients with COVID-19 were uploaded to the "Statistica" system of the company "StatSoft-Tibco". These acid-base state (ABS) samples consisted of 21 parameters, e.g. arteries: ABE_(art) — excess of bases, Ca^{2+}_(art) — calcium ion concentration, Cl^{-}_(art) – concentration of the chlorine ion, ctBil_(art) — bilirubin, $ctCO_2(B)$_(art) — total carbon dioxide content (calculated), $ctCO_2(P)$_(art) — total carbon dioxide content in the plasma (calculated), ctHb_(art) — reference hemoglobin level, ctO_2_(art) — the total concentration of oxygen in the blood including the concentration of oxygen dissolved in the plasma, Glu_(art) — glucose concentration, $HCO_3(P)$_(art) — plasma bicarbonate, Hct_(art) — hematocrit, Lac_(art) — lactate content, Na^{+}_(art) — sodium ion concentration, p50_(art) — the affinity of hemoglobin for oxygen, CO_2_(art) — partial pressure of oxygen, pH_(art) — acidity, O_2_(art) — partial pressure of oxygen, SBE_(art) — lack of bases, sO_2_(art) — oxygen saturation, K^{+}_(art) — potassium ion concentration, Osmolarity_(art) — arterial blood osmolarity.

A line-by-line deletion can destroy a noticeable part of the data if missing values are scattered throughout the data table or are located in several variables. And this is true in our case.

In such cases, a deeper analysis is usually made, using additional R-libraries for multiple substitution, such as "Mice". One can look at the relationship between the presence of missing values of one variable and the observed values of other variables and what happens in this case. The imputation for the parameters of the artery and vein was performed separately. 5 imputation sets were generated to select the best set.

The predicted mean matching method (PMM) was set as an imputation model for quantitative variables. PMM is a type of linear regression in which the imputation values calculated from the regression model are compared with the nearest observed values. The method of fully conditional specification of PMM — "predictive mean matching" was used due to the degeneracy of some regression parameters. The minimum and maximum values of the parameters of the 0-data set are used as restrictions on the maximum and minimum values for subsequent imputations (Table 1).

The data group the artery ABS. The table shows the minimum and maximum values of arterial blood parameters. They are the same for all five imputed data sets in order to avoid data distortion and prevent biases of statistical sets. These are the set values for imputation.

In all six data sets, including the initial imputed one, the entire set of 21 parameters for the artery and vein ABS were distributed according to the normal distribution law according to the Kolmogorov-Smirnov and Liliefors criteria with a probability less than 0.01.

The 5th group of the set of ABS for arteria with imputation is presented in Table 2. Similar sets of

Table 1. Average statistical parameters of the artery ABS in the group without imputation (0-data set)

Variable	Number of observations	Mean	Minimum	Maximum	Standard deviation
ABE_art	11871	0.4503	–31.0000	28.0000	5.10135
Ca^{2+}_art	11563	0.9676	0.2500	151.0000	1.40431
Cl^{-}_art	11541	109.9747	79.0000	149.0000	7.34768
ctBil_art	1601	26.8499	0.0000	265.0000	26.36076
$ctCO_2B$_art	11713	48.0726	7.4200	107.9000	12.01896
$ctCO_2P$_art	11513	55.2930	8.6000	121.3000	13.55506
ctHb_art	11885	107.7292	10.0000	195.0000	22.52998
ctO_2_art	11277	14.2235	1.8000	33.3000	2.98734
Glu_art	11647	8.9560	0.1000	37.0000	4.01030
HCO_3P_art	11894	24.7031	2.9000	59.1000	5.10861
Hct_art	2659	1.1126	0.0400	59.0000	5.10570
Lac_art	11684	2.5942	0.0000	30.0000	2.65705
Na_art	11895	142.3490	97.0000	195.0000	7.58373
p50_art	10533	25.7687	11.0000	262.4000	5.15931
pCO_2_art	11901	40.3392	10.0000	192.0000	13.14258
pH_art	11902	7.4116	6.7300	7.7600	0.09963
pO_2_art	11873	96.3650	18.0000	541.0000	41.05349
SBE_art	1013	1.7253	–27.2700	30.0000	7.60385
sO_2_art	11880	95.4498	10.5000	100.9000	6.50825
K^{+}_art	11861	4.1157	1.8000	19.6000	0.75089
Osmolarity_art	11317	293.0776	202.0000	402.0000	16.08119

Table 2. Average statistical parameters of the ABS of the artery in the imputation group 5 (5-data set)

Variable	Number of observations	Mean	Minimum	Maximum	Standard deviation
ABE_art	18658	0.4434	–31.0000	28.0000	5.18306
Ca^{2+}_art	18658	0.9885	0.2500	151.0000	2.20358
Cl^{-}_art	18658	109.9232	79.0000	149.0000	6.76504
ctBil_art	18658	35.1288	0.0000	265.0000	41.66691
$ctCO_2B$_art	18658	47.8398	7.4200	107.9000	9.98395
$ctCO_2P$_art	18658	54.6650	8.6000	121.3000	11.45273
ctHb_art	18658	107.2712	10.0000	195.0000	19.92765
ctO_2_art	18658	14.2043	1.8000	33.3000	2.60291
Glu_art	18658	9.0953	0.1000	37.0000	4.36174
HCO_3P_art	18658	24.6568	2.9000	59.1000	4.36005
Hct_art	18658	1.0042	0.0400	59.0000	4.95924
Lac_art	18658	2.5468	0.0000	30.0000	2.53684
Na_art	18658	142.3128	97.0000	195.0000	6.37804
p50_art	18658	25.0472	11.0000	262.4000	9.41725
pCO_2_art	18658	40.7309	10.0000	192.0000	15.32162
pH_art	18658	7.4134	6.7300	7.7600	0.14171
pO_2_art	18658	100.1013	18.0000	541.0000	46.94539
SBE_art	18658	0.0432	–27.2700	30.0000	5.36797
sO_2_art	18658	95.4232	10.5000	100.9000	6.30708
K^{+}_art	18658	4.1217	1.8000	19.6000	0.82043
Osmolarity_art	18658	293.2238	202.0000	402.0000	13.34610

data imputation were obtained for the models of the ABS vein.

It is also noteworthy that all significantly different parameters have smaller average values compared to the initial indicators and smaller standard deviations (variances), which indicates that at least the imputed values do not fluctuate more significantly in the vicinity of their average values, compared with the initial sample.

To assess the degree of differences between the original array and the imputed data, we performed a discriminant analysis between these samples (sets of imputations). Canonical discriminant functions for imputed artery data sets are obtained (Table 3).

For the analysis, the first 4 of the canonical discriminant functions were used, which 100 % describe the variances and the first of which "explains" the process of "scattering" by 96.3 %, and the second by 2.1 % (in total 98.4 %). The first two canonical discriminant functions for the 1st data set (without imputation) and the fifth set of imputed data are shown in Fig. 1.

According to Table 3, the results of classification and comparison of forecasts for belonging to the implanted artery data were obtained and showed that 21.9 % of the initial grouped observations were classified correctly. This means that only 22 % of the imputed observations differ from each other, which implies that the differences between the imputed sets of arterial blood parameters are not significant. This fact suggests that the imputed sets do not differ from each other, but they allow for the problems of system models to use correctly (in the sense of without omissions) the criteria functions of the ABS, which will be discussed below.

The graphical results of comparing the canonical discriminant functions (for two canonical functions that explain 98.4 % of the variance of the artery parameters in total) are presented in Fig. 1, in which the practical equivalence of the initial and the fifth imputed artery data sets is clearly demonstrated.

Canonical discriminant functions for imputed vein data sets are obtained in the same way (Table 4).

Table 3. Canonical discriminant functions of the ABS artery and their eigenvalues

Function	Eigenvalue	Variances, %	Total, %	Canonical correlation
1	0.123	96.3	96.3	0.331
2	0.003	2.1	98.4	0.052
3	0.002	1.3	99.7	0.041
4	0.000	0.3	100.0	0.021

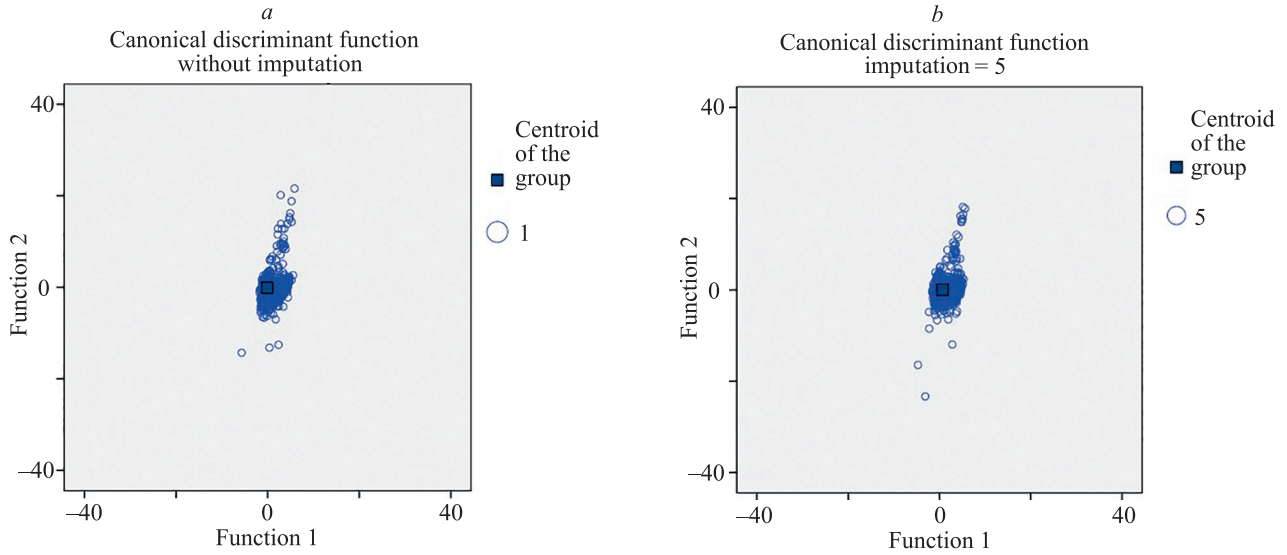

Fig. 1. Comparison of canonical discriminant functions (centroids of groups) of the ABS artery for data sets without imputation (*a*) and in the 5th group of imputation (*b*)

Table 4. Canonical discriminant functions of the ABS vein and their eigenvalues

Function	Eigenvalue	Variances, %	Total, %	Canonical correlation
1	0.101	93.8	93.8	0.303
2	0.004	3.7	97.5	0.063
3	0.002	1.8	99.3	0.043
4	0.001	0.7	100.0	0.028

For the analysis, the first 4 of the canonical discriminant functions were used, which 100 % describe the variances and the first of which "explains" the process of "scattering" by 93.8 %, and the second by 3.7 % (in total 97.5 %). The first two canonical discriminant functions for the 1st data set (without imputation) and the fifth set of imputed vein data are shown in Fig. 2.

The results of classification and comparison of forecasts for belonging to the imputed data of the vein are obtained, from which it follows that 25.9 % of the initial grouped observations are classified correctly. This means that only 26 % of the implanted observations differ from each other, which means that the differences between the implanted sets of venous blood parameters are not significant.

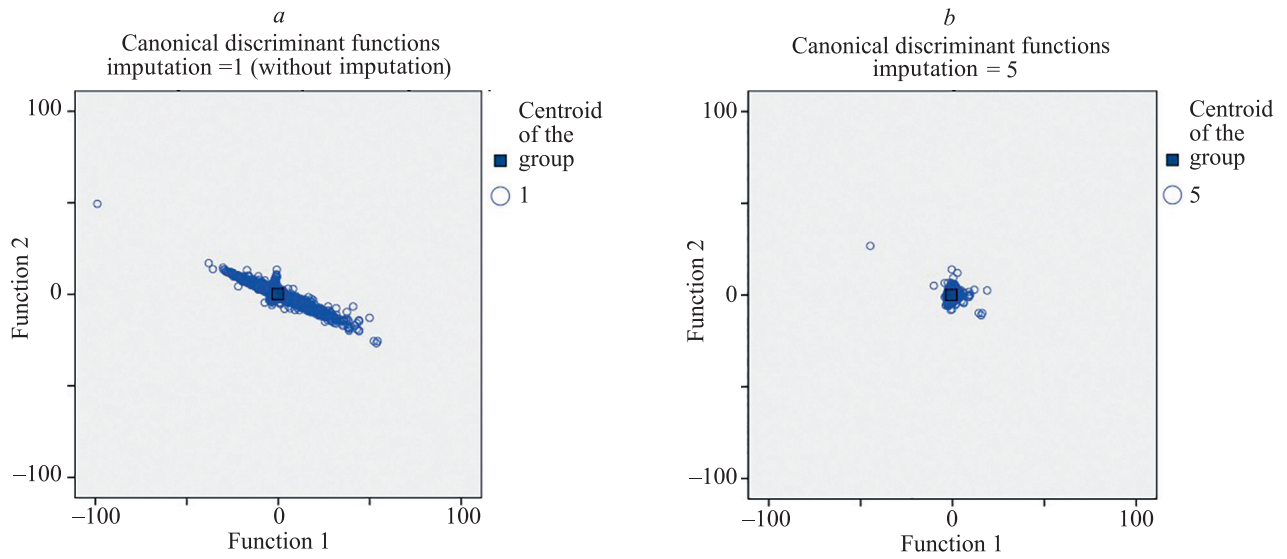

Fig. 2. Comparison of canonical discriminant functions (centroids of groups) of braid veins for data sets without imputation (*a*) and in the 5th group of imputation (*b*)

Table 5. Hotelling T^2 statistics for comparing the parameters of ABS in the artery and vein of sample groups: “Died”, “Discharged”, “Transferred to another medical facility”

Vessel parameter	Comparison groups	Hotelling T^2
Art	Died – Discharged	1220.570
Art	Died – Transferred to another medical facility	1742.560
Art	Discharged – Transferred to another medical facility	741.517
Vein	Died – Discharged	1696.840
Vein	Died – Transferred to another medical facility	496.160
Vein	Discharged – Transferred to another medical facility	351.272

The graphical results of comparing the canonical discriminant functions (for two canonical functions that explain 97.5 % of the variances of the vein parameters in total) are presented in Fig. 2. The practical equivalence of the original and the fifth imputed data sets is clearly demonstrated.

Let us move on to a multidimensional comparison based on the statistics of Hotelling T^2 of imputed data between 3 groups of patients (Table 5): “discharged”, “died”, “transferred to another medical facility”. A significance level $p = 0.0000$ for F-statistics was obtained for all comparison groups.

From Table 5, presented above, it follows that the 21 ABS parameters in both artery and vein in a multidimensional space differ in pairs for 3 “states” according to the multidimensional statistics of Hotelling T^2. This remarkable fact confirms, firstly, the significant difference between these conditions in the outcomes of treatment of patients with COVID-19, secondly, the adequacy of the applied multidimensional Hotelling T^2 statistics, and thirdly, the correctness of the initial data imputation, along with the results of the discriminant analysis with its canonical discriminant functions and centroids.

Dynamic system statistical analysis of ABS parameters in DGP

Now we present the results of a systematic dynamic analysis of the ABS for each individual patient with COVID-19. System analysis assumes a fully filled data matrix. Hence, the reasons for the above data imputation are clear.

The average statistical indicators of the decimal logarithms of the CF ABS were calculated on an IBM-compatible PC with a clock frequency of 2.5 GHz of the Intel i5-2400S processor for 7 hours. The reliability of the differences in the averages and variances for survivors, deceased, and transferred to other medical institutions of patients is calculated; the results are shown in Fig. 3.

It follows from Fig. 3 that both the mean and variance values of the logarithm of the CF of the ABS in patients with COVID-19 differ for the artery and vein according to the outcomes in the groups “discharged”, “died” and “transferred to other medical facilities” significantly. This fact is confirmed by the graphs of the average values of this indicator by groups. The results are noteworthy: despite the very insignificant numerical differences between the averages and variances, they are nevertheless statistically significant. In addition, these are decimal logarithms, and they are the power of the number that needs to be raised to get the true value of the CF of the ABS, which in turn reflects the “level of correlation” within the ABS system of a particular patient. Moreover, since the logarithms have a value greater than 3 in our case, the indicators of CF ABS have an order of magnitude from 1000 to 10 000, and the differences in the average logarithms for the deceased and those discharged from the clinic are even on near 0.12, they can have a statistically significant control effect on the ABS system.

According to the data in Fig. 3, it can be seen that all the compared groups differ significantly in mean values and variances, except for the absence of differences in the average values of the logarithm of the CF ABS vein between the groups “discharged” –“transferred to another medical facility” (Fig. 3, *f*).

Thus, from Fig. 3, a pairwise difference in the logarithms of the CF ABS for all groups of states (“died”, “discharged”, “transferred to another medical facility”) follows, which emphasizes the diagnostic value of the system indicator under consideration.

We will demonstrate how the decimal logarithm of the system criterion function of the ABS behaves for individual groups of patients: “died” (Fig. 4), “discharged” (Fig. 5), “transferred to another medical facility” (Fig. 6).

Discussion

The results of the studies revealed a certain statistically significant determinism in the response of ABS to Covid infection (DGP) from the standpoint of dividing patients into groups: “discharged”, “died”, and “transferred to another medical facility”. There is probably a certain uncertainty in the outcome of the disease, depending on the initial state of the ABS at the time of contact of the coronavirus with the human body. This is indicated by the “levels – groups” of outcomes. This paper presents the results of studies for men and women at all ages. The results for different age categories for men and women may likely differ. Subsequent studies may reveal the peculiarities of the response of gender and age groups of people to contact with the coronavirus.

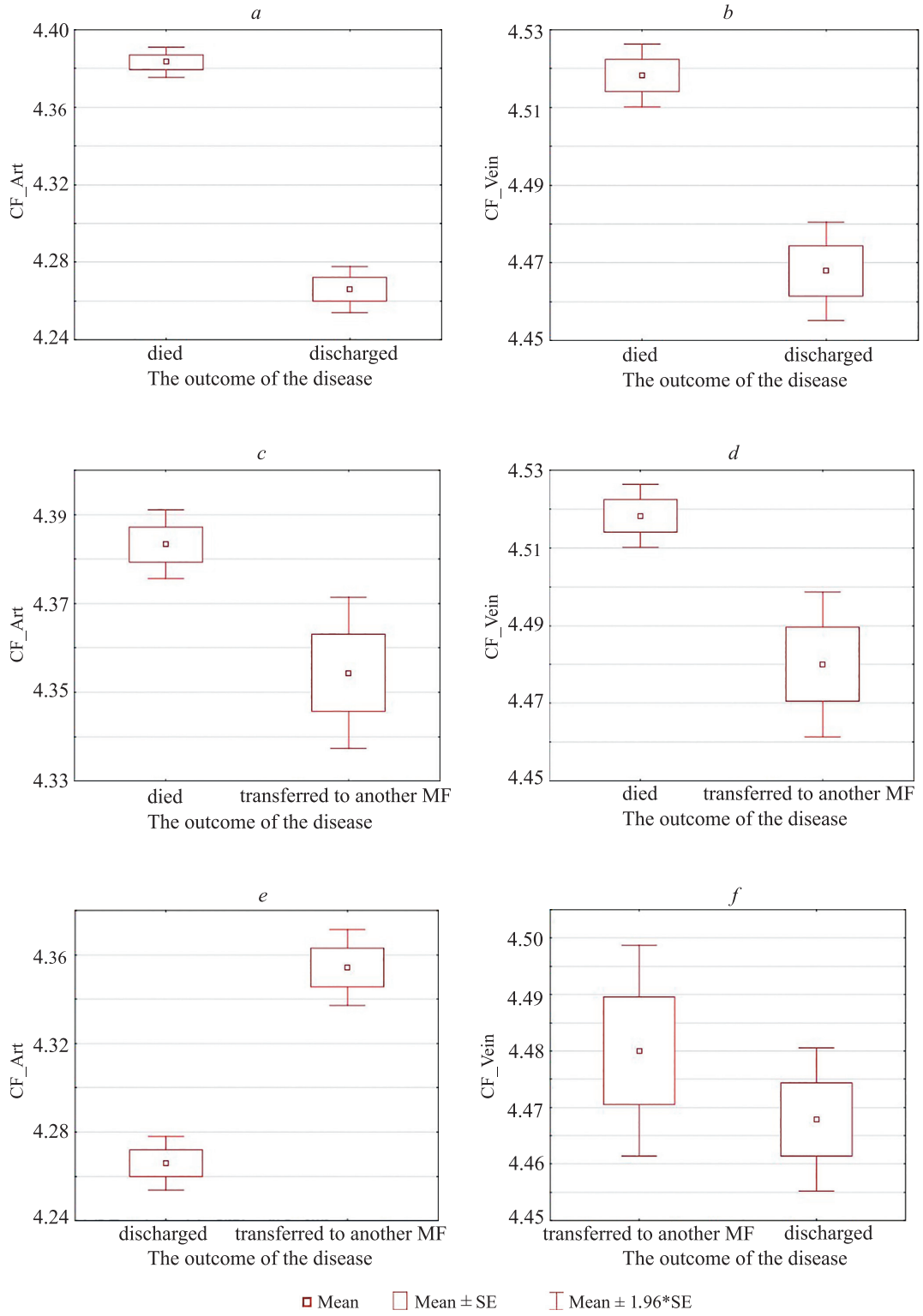

Fig. 3. Graphs of pairwise differences in the average values of the logarithms of the CF ABS of the arteries and veins in groups: “died”, “discharged” (*a*) and (*b*); “died”, “transferred to another medical facility” (*c*) and (*d*); “discharged”, “transferred to another medical facility” (*e*) and (*f*)

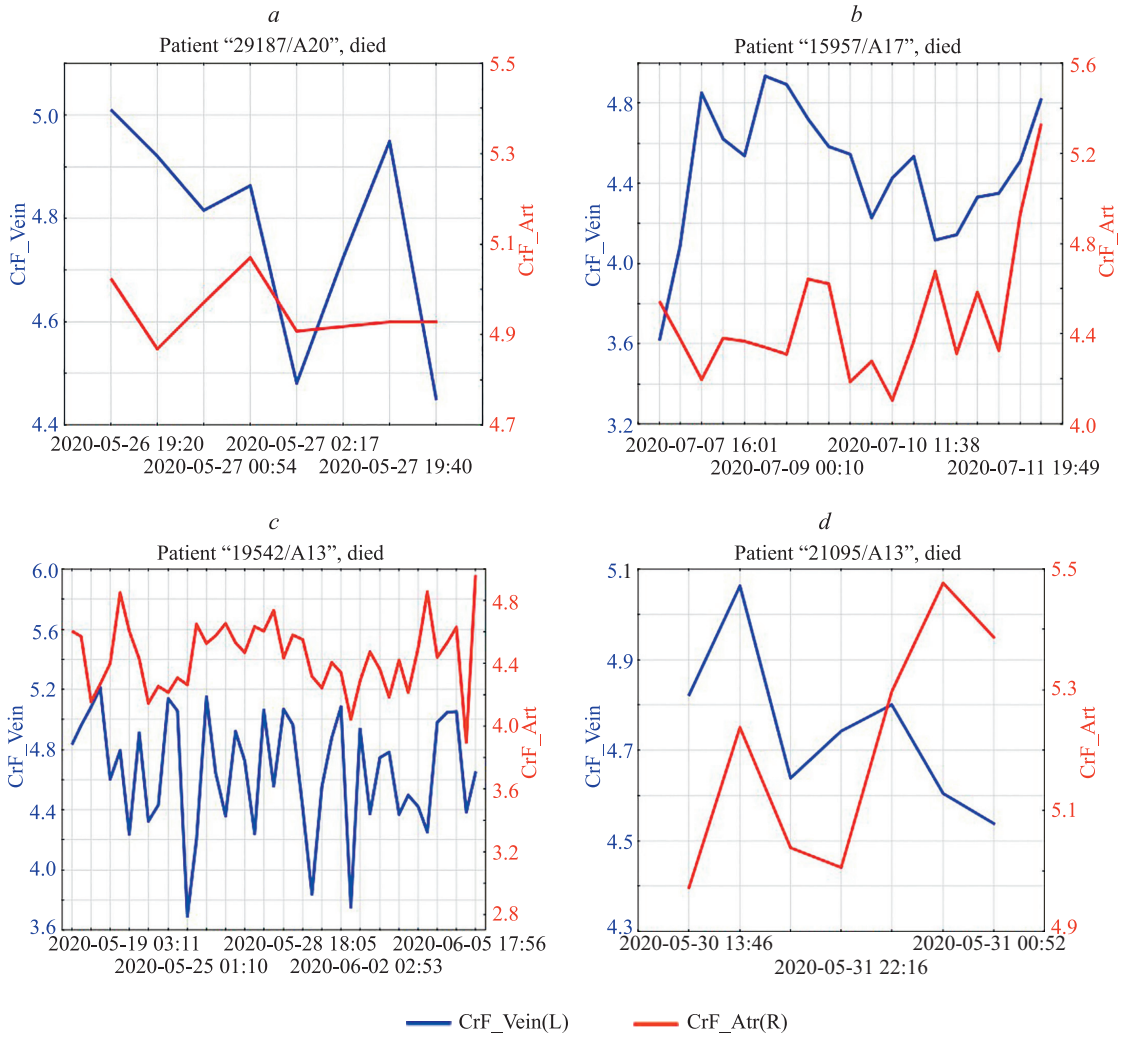

Fig. 4. Examples of the dynamics of logarithms of CF ABS by patient groups: “Died”: severe patient (fatal outcome, arterial and vein indicators are almost synchronous) (*a*) and (*b*); fatal patient (there is a tendency to synchronicity of arterial and vein indicators, at the final stage of the trend (the last third), the opposite phase of the processes is visible) (*c*) and (*d*)

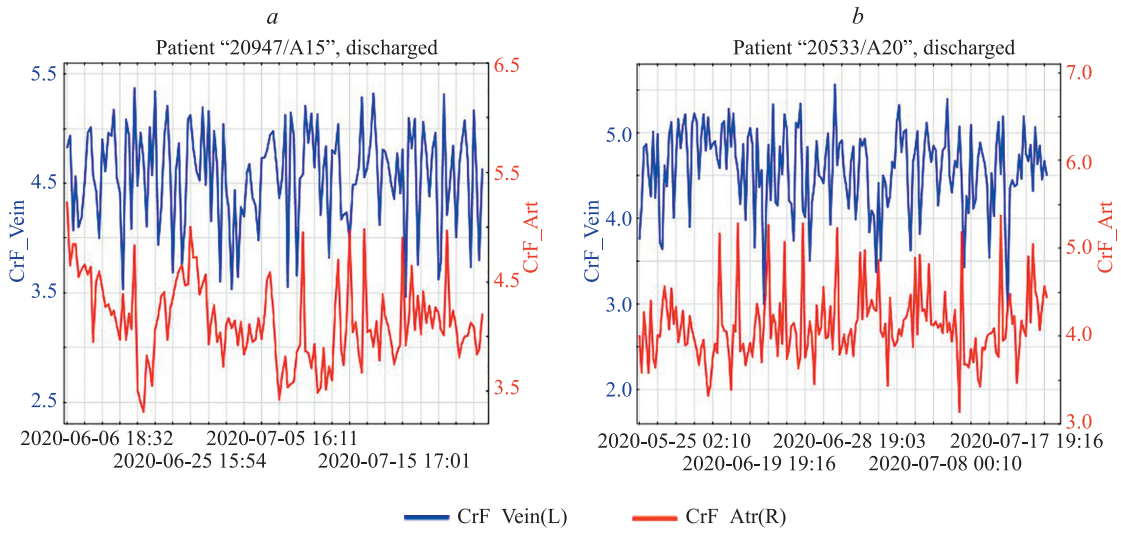

Fig. 5. Dynamics of logarithms of CF ABS by patient groups: “Discharged”. In both cases, the logarithms demonstrate relative stability (*a*) and (*b*)

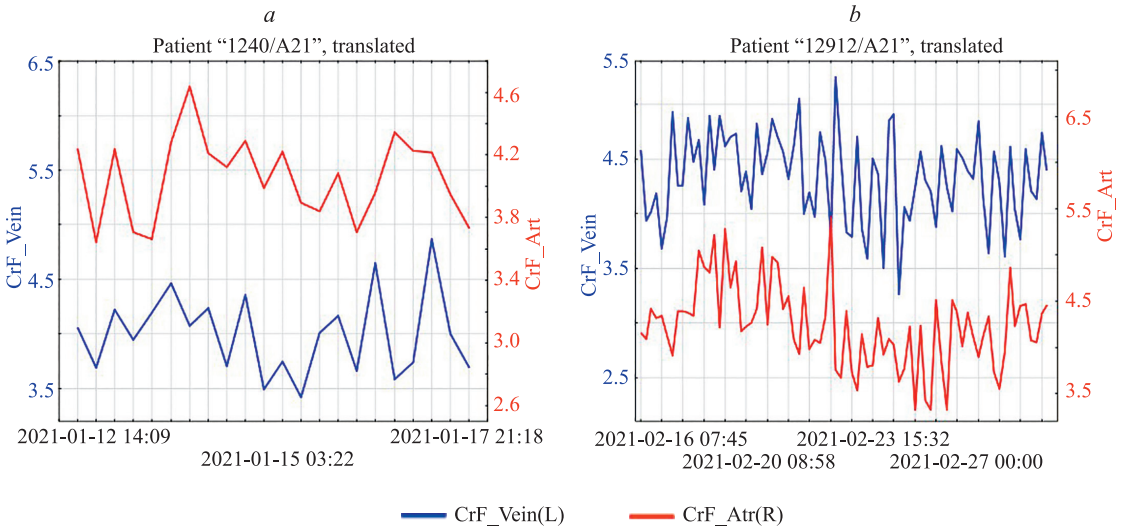

Fig. 6. Examples of the dynamics of logarithms of CF ABS by patient groups: "Transferred to another medical facility". Logarithms demonstrate a long-term trend and synchronicity (*a*) and (*b*)

The clinical implications of this work are that research doctors receive a new analytical tool for evaluating the multidimensional ABS of DGP in the form of a criterion function, by which it is possible to determine the dynamics of a particular patient's system. The social consequences of this scientific work show that when treating and evaluating patients with COVID-19, it is necessary to pay attention to the prognostic value of the dynamics of CF ABS, which can indicate the possibility of a fatal outcome in a certain category of patients.

Conclusion

The work shows:

— Data imputation (substitution) significantly increases the volume of the verified sample.

— The substituted data make it possible to carry out a systematic statistical assessment of the totality of the parameters of the organism based on the calculation of the logarithms of the criteria functions of the ABS.

— The logarithm of CF ABS, being a systematic statistical assessment, allows us to distinguish in DGP by outcomes in three groups reliably: "discharged", "died", and "transferred to another medical facility".

References

1. Anokhin P.K. Theory of a functional system. *Uspekhi Fiziologicheskikh Nauk*, 1970, vol. 1, no. 1, pp. 19–54. (in Russian)
2. Haken G. Information and Self-Organization: *A Macroscopic Approach to Complex Systems*. Springer-Verlag, 1988, 188 p.
3. Lushnov A.M., Lushnov M.S. *Medical Information Systems: Multidimensional Analysis of Medical and Environmental Data*. Saint Petersburg, Helikon Plus, 2013, 458 p. (in Russian)
4. Kurapeev D.I., Lushnov M.S., Osipov V.Yu., Vodyaho A.I., Zhukova N.A. Synthesis of integral models of system dynamics of an Acid-Base State (ABS) of patients at operative measures. *Acta Scientific Medical Sciences*, 2019, vol. 3, no. 3, pp. 16–29.
5. Kurapeev D.I., Lushnov M.S., Zhukova N.A. Comparison of integral models of systemic dynamics of Acid-Base State of venous and arterial circulating blood in patients with surgical interventions. *Acta Scientific Medical Sciences*, 2019, vol. 3, no. 6, pp. 38–48.
6. Lebedev S., Zhukova N., Vodyaho A., Kurapeev D., Lushnov M. An ontology-driven toolset for fast prototyping of medical data processing systems. *International Journal of Biology and Biomedical Engineering*, 2017, vol. 11, pp. 135–142.
7. Narenda P.M., Fukunaga K. A branch and bound algorithm for feature subset selection. *IEEE Transactions Computers*, 1977, vol. C-26, no. 9, pp. 917–922. https://doi.org/10.1109/TC.1977.1674939
8. Ridout M.S. An improved branch and bound algorithm for feature subset-selection. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1988, vol. 37, no. 1, pp. 139–147. https://doi.org/10.2307/2347512
9. Bech C.N., Brabrand M., Mikkelsen S., Lassen A. Risk factors associated with short term mortality changes over time, after arrival to the emergency department. *Scandinavian Journal of Trauma*,

Литература

1. Анохин П.К. Теория функциональной системы // Успехи физиологических наук. 1970. Т. 1. № 1. С. 19–54.
2. Хакен Г. Информация и самоорганизация: Макроскопический подход к сложным системам / пер. с англ. М.: Мир, 1991. 240 с.
3. Лушнов А.М., Лушнов М.С. Медицинские информационные системы: многомерный анализ медицинских и экологических данных. СПб.: Геликон Плюс, 2013. 458 с.
4. Kurapeev D.I., Lushnov M.S., Osipov V.Yu., Vodyaho A.I., Zhukova N.A. Synthesis of integral models of system dynamics of an Acid-Base State (ABS) of patients at operative measures // Acta Scientific Medical Sciences. 2019. V. 3. N. 3. P. 16–29.
5. Kurapeev D.I., Lushnov M.S., Zhukova N.A. Comparison of integral models of systemic dynamics of Acid-Base State of venous and arterial circulating blood in patients with surgical interventions // Acta Scientific Medical Sciences. 2019. V. 3. N 6. P. 38–48.
6. Lebedev S., Zhukova N., Vodyaho A., Kurapeev D., Lushnov M. An ontology-driven toolset for fast prototyping of medical data processing systems // International Journal of Biology and Biomedical Engineering. 2017. V. 11. P. 135–142.
7. Narenda P.M., Fukunaga K. A branch and bound algorithm for feature subset selection // IEEE Transactions Computers. 1977. V. C-26. N 9. P. 917–922. https://doi.org/10.1109/TC.1977.1674939
8. Ridout M.S. An improved branch and bound algorithm for feature subset-selection // Journal of the Royal Statistical Society. Series C (Applied Statistics). 1988. V. 37. N 1. P. 139–147. https://doi.org/10.2307/2347512
9. Bech C.N., Brabrand M., Mikkelsen S., Lassen A. Risk factors associated with short term mortality changes over time, after arrival to the emergency department // Scandinavian Journal of Trauma,

Resuscitation and Emergency Medicine, 2018, vol. 26, no. 1, pp. 29. https://doi.org/10.1186/s13049-018-0493-2
10. Fuchs P.A., Del Junco D.J., Fox E.E., Holcomb J.B., Rahbar M.H., Wade C.A., Alarcon L.H., Brasel K.J., Bulger E.M., Cohen M.J., Myers J.G., Muskat P., Phelan H.A., Schreiber M.A., Cotton B.A. Purposeful variable selection and stratification to impute missing Focused Assessment with Sonography for Trauma data in trauma research. *Journal of Trauma and Acute Care Surgery*, 2013, vol. 75, no. 1 (suppl. 1), pp. S75–S81. https://doi.org/10.1097/TA.0b013e31828fa51c
11. Mantrova A.I. How can omissions in medical data affect the results of research? *Scientific Review. Biological Sciences*, 2019, no. 2, pp. 5–9. (in Russian)
12. Lefering R., Huber-Wagner S., Nienaber U., Maegele M., Bouillon B. Update of the trauma risk adjustment model of the TraumaRegister DGU™: the Revised Injury Severity Classification, version II. *Critical Care*, 2014, vol. 18, no. 5, pp. 476. https://doi.org/10.1186/s13054-014-0476-2
13. Seleno N., Vogel J., Liao M., Hopkins E., Byyny R., Moore E., Gravitz C., Haukoos J. Denver trauma organ failure score outperforms traditional methods of risk stratification in trauma. *Academic Emergency Medicine*, 2012, vol. 19, suppl. 1, pp. S144.
14. Zolin P.P. Statistical processing of censored samples in the study of extreme and terminal states. *Pathogenesis, clinical findings and therapy of extreme and terminal states: Proceedings of scientific and practical conference*. Omsk, 1998, pp. 39–43. (in Russian)
15. Fabian-Jessing B.K., Vallentin M.F., Secher N., Hansen F.B., Dezfulian C., Granfeldt A., Andersen L.W. Animal models of cardiac arrest: A systematic review of bias and reporting. *Resuscitation*, 2018, vol. 125, pp. 16–21. https://doi.org/10.1016/j.resuscitation.2018.01.047
16. Krantz M.J., Kaul S. The ATLAS ACS 2–TIMI 51 trial and the burden of missing data: (Anti-Xa Therapy to Lower Cardiovascular Events in Addition to Standard Therapy in Subjects With Acute Coronary Syndrome ACS 2–Thrombolysis In Myocardial Infarction 51). *Journal of the American College of Cardiology*, 2013, vol. 62, no. 9, pp. 777–781. https://doi.org/10.1016/j.jacc.2013.05.024
17. Little R.J., D'Agostino R., Cohen M.L., Dickersin K., Emerson S.S., Farrar J.T., Frangakis C., Hogan J.W., Molenberghs G., Murphy S.A., Neaton J.D., Rotnitzky A., Scharfstein D., Shih W.J., Siegel J.P., Stern H. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 2012, vol. 367, pp. 1355–1360. https://doi.org/10.1056/NEJMsr1203730
18. *National Research Council. The prevention and treatment of missing data in clinical trials*. Washington, DC, National Academies Press, 2010, 162 p.
19. Fabrykant M.S. Model-oriented approach to missing values: Multiple imputation in multilevel regression using R (on the example of analyzing survey data). *Sociology: methodology, methods, mathematical modeling (Sociology:4M)*, 2015, no. 41, pp. 7–29. (in Russian)
20. Van Buuren S. *Flexible Imputation of Missing Data*. CRC Press, 2012, 342 p. https://doi.org/10.1201/b11826
21. Kulikova K.Yu. Statistical analysis of medical data with missing values. *Management processes and sustainability*, 2014, vol. 1, no. 1, pp. 253–258. (in Russian)
22. Witkiewitz K., Falk D.E., Kranzler H.R., Litten R.Z., Hallgren K.A., O'Malley S.S., Anton R.F. Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome measure when participants drop out from treatment in alcohol clinical trials. *Alcoholism: Clinical and Experimental Research*, 2014, vol. 38, no. 11, pp. 2826–2834. https://doi.org/10.1111/acer.12543
23. Parvez B., Shah A., Muhammad R., Shoemaker M.B., Graves A.J., Heckbert S.R., Xu H., Ellinor P.T., Benjamin E.J., Alonso A., Shintani A.K., Roden D., Darbar D. Replication of a risk prediction model for ambulatory incident atrial fibrillation using electronic medical record. *Circulation*, 2012, vol. 126, no. 21, meeting abstract 18578.
24. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of Translational Medicine*, 2016, vol. 4, no. 2, pp. 30. https://doi.org/10.3978/j.issn.2305-5839.2015.12.63
25. Aladyshkina A.S., Lakshina V.V., Leonova L.A., Maksimov A.G. Working with data on population health: imputation. *Social aspects of population health*, 2020, vol. 66, no. 1, pp. 12. (in Russian). https://doi.org/10.21045/2071-5021-2020-66-1-12

Resuscitation and Emergency Medicine. 2018. V. 26. N 1. P. 29. https://doi.org/10.1186/s13049-018-0493-2
10. Fuchs P.A., Del Junco D.J., Fox E.E., Holcomb J.B., Rahbar M.H., Wade C.A., Alarcon L.H., Brasel K.J., Bulger E.M., Cohen M.J., Myers J.G., Muskat P., Phelan H.A., Schreiber M.A., Cotton B.A. Purposeful variable selection and stratification to impute missing Focused Assessment with Sonography for Trauma data in trauma research // Journal of Trauma and Acute Care Surgery. 2013. V. 75. N 1 (Suppl. 1). P. S75–S81. https://doi.org/10.1097/TA.0b013e31828fa51c
11. Мантрова А.И. Как пропуски в медицинских данных могут влиять на результаты исследований? // Научное обозрение. Биологические науки. 2019. № 2. С. 5–9.
12. Lefering R., Huber-Wagner S., Nienaber U., Maegele M., Bouillon B. Update of the trauma risk adjustment model of the TraumaRegister DGU™: the Revised Injury Severity Classification, version II // Critical Care. 2014. V. 18. N 5. P. 476. https://doi.org/10.1186/s13054-014-0476-2
13. Seleno N., Vogel J., Liao M., Hopkins E., Byyny R., Moore E., Gravitz C., Haukoos J. Denver trauma organ failure score outperforms traditional methods of risk stratification in trauma // Academic Emergency Medicine. 2012. V. 19. Suppl. 1. P. S144.
14. Золин П.П. Статистическая обработка цензурированных выборок при изучении экстремальных и терминальных состояний // Патогенез, клиника и терапия экстремальных и терминальных состояний: Материалы научно-практической конференции, г. Омск, 21 октября 1998 г. Омск: Омская государственная медицинская академия, 1998. С. 39–43.
15. Fabian-Jessing B.K., Vallentin M.F., Secher N., Hansen F.B., Dezfulian C., Granfeldt A., Andersen L.W. Animal models of cardiac arrest: A systematic review of bias and reporting // Resuscitation. 2018. V. 125. P. 16–21. https://doi.org/10.1016/j.resuscitation.2018.01.047
16. Krantz M.J., Kaul S. The ATLAS ACS 2–TIMI 51 trial and the burden of missing data: (Anti-Xa Therapy to Lower Cardiovascular Events in Addition to Standard Therapy in Subjects With Acute Coronary Syndrome ACS 2–Thrombolysis In Myocardial Infarction 51) // Journal of the American College of Cardiology. 2013. V. 62. N 9. P. 777–781. https://doi.org/10.1016/j.jacc.2013.05.024
17. Little R.J., D'Agostino R., Cohen M.L., Dickersin K., Emerson S.S., Farrar J.T., Frangakis C., Hogan J.W., Molenberghs G., Murphy S.A., Neaton J.D., Rotnitzky A., Scharfstein D., Shih W.J., Siegel J.P., Stern H. The prevention and treatment of missing data in clinical trials // New England Journal of Medicine. 2012. V. 367. P. 1355–1360. https://doi.org/10.1056/NEJMsr1203730
18. National Research Council. The prevention and treatment of missing data in clinical trials. Washington, DC: National Academies Press, 2010. 162 p.
19. Фабрикант М.С. Модель-ориентированный подход к отсутствующим значениям: множественная импутация в многоуровневой регрессии посредством R (на примере анализа опросных данных) // Социология: методология, методы, математическое моделирование (Социология:4М). 2015. № 41. С. 7–29.
20. Van Buuren S. Flexible Imputation of Missing Data. CRC Press, 2012. 342 p. https://doi.org/10.1201/b11826
21. Куликова К.Ю. Статистический анализ медицинских данных при наличии пропусков // Процессы управления и устойчивость. 2014. Т. 1. № 1. С. 253–258.
22. Witkiewitz K., Falk D.E., Kranzler H.R., Litten R.Z., Hallgren K.A., O'Malley S.S., Anton R.F. Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome measure when participants drop out from treatment in alcohol clinical trials // Alcoholism: Clinical and Experimental Research. 2014. V. 38. N 11. P. 2826–2834. https://doi.org/10.1111/acer.12543
23. Parvez B., Shah A., Muhammad R., Shoemaker M.B., Graves A.J., Heckbert S.R., Xu H., Ellinor P.T., Benjamin E.J., Alonso A., Shintani A.K., Roden D., Darbar D. Replication of a risk prediction model for ambulatory incident atrial fibrillation using electronic medical record // Circulation. 2012. V. 126. N 21. Meeting Abstract 18578.
24. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package // Annals of Translational Medicine. 2016. V. 4. N 2. P. 30. https://doi.org/10.3978/j.issn.2305-5839.2015.12.63
25. Аладышкина А.С., Лакшина В.В., Леонова Л.А., Максимов А.Г. Особенности работы с данными, характеризующими здоровье

26. Little R.J.A., Rubin D.B. *Statistical Analysis with Missing Data*. John Wiley & Sons, 2014, 408 p.
27. Seitz C., Lanius V., Lippert S., Gerlinger C., Haberland C., Oehmke F., Tinneberg H.-R. Patterns of missing data in the use of the endometriosis symptom diary. *BMC Women's Health*, 2018, vol. 18, no. 1, pp. 88. https://doi.org/10.1186/s12905-018-0578-0

населения: заполнение пропусков в данных // Социальные аспекты здоровья населения. 2020. Т. 66. № 1. С. 12. https://doi.org/10.21045/2071-5021-2020-66-1-12
26. Little R.J.A., Rubin D.B. Statistical Analysis with Missing Data. John Wiley & Sons, 2014. 408 p.
27. Seitz C., Lanius V., Lippert S., Gerlinger C., Haberland C., Oehmke F., Tinneberg H.-R. Patterns of missing data in the use of the endometriosis symptom diary // BMC Women's Health. 2018. V. 18. N 1. P. 88. https://doi.org/10.1186/s12905-018-0578-0

Authors

Dmitry I. Kurapeev — PhD, Chief Information Officer, Almazov National Medical Research Centre, Saint Petersburg, 197341, Russian Federation, SC 57225231263, https://orcid.org/0000-0002-2190-1495, dkurapeev@gmail.com

Mikhail S. Lushnov — D.Sc., Methodologist, Almazov National Medical Research Centre, Saint Petersburg, 197341, Russian Federation, SC 57190125064, https://orcid.org/0000-0002-9683-1858, Lushnov_ms@almazovcentre.ru

Tianxing Man — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57209975532, https://orcid.org/0000-0003-2187-1641, mantx626@gmail.com

Natalia A. Zhukova — PhD, Associate Professor, Leading Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation; Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 56406142300, https://orcid.org/0000-0001-5877-4461, nazhukova@mail.ru

Авторы

Курапеев Дмитрий Ильич — кандидат медицинских наук, заместитель генерального директора, Национальный медицинский исследовательский центр имени В.А. Алмазова, Санкт-Петербург, 197341, Российская Федерация, SC 57225231263, https://orcid.org/0000-0002-2190-1495, dkurapeev@gmail.com

Лушнов Михаил Степанович — доктор медицинских наук, врач-методист, Национальный медицинский исследовательский центр имени В.А. Алмазова, Санкт-Петербург, 197341, Российская Федерация, SC 57190125064, https://orcid.org/0000-0002-9683-1858, Lushnov_ms@almazovcentre.ru

Ман Тяньсин — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57209975532, https://orcid.org/0000-0003-2187-1641, mantx626@gmail.com

Жукова Наталия Александровна — кандидат технических наук, доцент, ведущий научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация; доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 56406142300, https://orcid.org/0000-0001-5877-4461, nazhukova@mail.ru

Received 17.10.2021
Approved after reviewing 07.12.2021
Accepted 28.01.2022

Статья поступила в редакцию 17.10.2021
Одобрена после рецензирования 07.12.2021
Принята к печати 28.01.2022