

ЛЕКЦИЯ

LECTURE

УДК [616.12-089]

<https://doi.org/10.17816/MAJ17737>

SAS ENTERPRISE GUIDE 6.1 ДЛЯ ВРАЧЕЙ: КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Н.С. Буненков, Г.Ф. Буненкова, В.В. Комок, О.А. Гриненко, А.С. Немков

Федеральное государственное бюджетное образовательное учреждение высшего образования «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова»
Министерства здравоохранения Российской Федерации, Санкт-Петербург

Для цитирования: Буненков Н.С., Буненкова Г.Ф., Комок В.В., и др. SAS Enterprise Guide 6.1 для врачей: корреляционный анализ // Медицинский академический журнал. – 2020. – Т. 20. – № 1. – С. 51–56. <https://doi.org/10.17816/MAJ17737>

Поступила: 14.11.2019

Одобрена: 13.01.2020

Принята: 02.03.2020

Цель — разработать алгоритм обработки базы данных проспективного нерандомизированного исследования Assessment of Myocardial Ischemic-Reperfusion Injury During Off- and On-Pump CABG (AMIRI-CABG, ClinicalTrials.gov Identifier: NCT03050489) в программном пакете SAS Enterprise Guide 6.1 для определения наличия и оценки силы корреляции между исследуемыми переменными.

Материалы и методы. В проспективное нерандомизированное исследование AMIRI-CABG (ClinicalTrials.gov Identifier: NCT03050489) в ПСПбГМУ им. И.П. Павлова с 2016 по 2019 г. включено 336 пациентов. Создана база данных с результатами клинических, лабораторных и инструментальных исследований. Статистическая обработка данных выполнена с использованием лицензионного программного обеспечения SAS Enterprise Guide 6.1. Корреляционный анализ проводили с применением коэффициентов Пирсона и Спирмена.

Результаты. Разработан алгоритм обработки данных проспективного нерандомизированного исследования AMIRI-CABG (ClinicalTrials.gov Identifier: NCT03050489), позволяющий выполнить корреляционный анализ. Данный алгоритм может быть использован врачами различных специальностей и научными сотрудниками для обработки результатов научных или клинических исследований.

Заключение. С помощью разработанного алгоритма обработки базы данных научного исследования и программного пакета SAS Enterprise Guide 6.1 врачи и научные сотрудники смогут упростить и ускорить корреляционный анализ результатов исследований.

Ключевые слова: SAS Enterprise Guide; статистическая обработка данных; статистика; клинические исследования; научные исследования; корреляционный анализ; коэффициент корреляции Пирсона; коэффициент корреляции Спирмена.

SAS ENTERPRISE GUIDE 6.1 FOR PHYSICIANS: CORRELATION ANALYSIS

N.S. Bunenkov, G.F. Bunenkova, V.V. Komok, A.A. Grinenko, A.S. Nemkov

Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russia

For citation: Bunenkov NS, Bunenkova GF, Komok VV, et al. SAS Enterprise Guide 6.1 for physicians: correlation analysis. *Medical Academic Journal*. 2020;20(1):51-56. <https://doi.org/10.17816/MAJ17737>

Received: November 14, 2019

Revised: January 13, 2020

Accepted: March 2, 2020

Objective: to develop algorithm of correlation analysis of prospective non-randomized clinical trial AMIRI-CABG (ClinicalTrials.gov Identifier: NCT03050489) data using SAS Enterprise Guide 6.1.

Materials and methods. Data collection was performed according prospective non-randomized clinical trial AMIRI-CABG (ClinicalTrials.gov Identifier: NCT03050489) in Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russia between 2016–2019 years with 336 patients. There is database with clinical, laboratory and instrumental data. Correlation analysis was performed with SAS Enterprise Guide 6.1.

Results. There was developed algorithm of correlation analysis data of prospective non-randomized clinical trial AMIRI-CABG (ClinicalTrials.gov Identifier: NCT03050489). This algorithm could be useful for physicians and researchers for data analysis.

Conclusion. Presented algorithm of correlation analysis could make easier and improve efficient data analysis with SAS Enterprise Guide 6.1.

Keywords: SAS Enterprise Guide 6.1; statistical analysis; clinical trials; correlation analysis; Pearson coefficient correlation; Spearman coefficient correlation.

Введение

Одной из целей научных исследований зачастую является установление зависимости между изучаемыми параметрами, например маркером в плазме крови и концентрацией препарата. Распространено мнение, что для выявления зависимости между двумя переменными необходимо оценить корреляцию, и если она существует, то говорят, что корреляция есть. Нередко слово «зависимость» заменяют словом «корреляция», и можно услышать вопрос: «Наблюдалась ли корреляция?» Однако наличие корреляции не подразумевает прямую причинно-следственную связь, так же как и отсутствие корреляции не исключает наличия связи между двумя переменными, в том числе причинно-следственной [3]. В ходе научного исследования в кардиохирургическом центре ПСПбГМУ им. акад. И.П. Павлова возник вопрос: есть ли взаимосвязь между временем пережатия аорты при операциях коронарного шунтирования и уровнем тропонина I после операции?

В нашем примере мы взяли для анализа следующие переменные: TnIEndOp — уровень тропонина I к концу операции, TnI1 — уровень тропонина I на первые сутки после операции коронарного шунтирования, AoClamp — время пережатия аорты в минутах. Было сформировано три группы пациентов: первая группа — коронарное шунтирование без искусственного кровообращения, вторая группа — коронарное шунтирование с искусственным кровообращением, третья группа — коронарное шунтирование на параллельном искусственном кровообращении.

Цель данной работы состояла в написании алгоритма обработки базы данных проспективного нерандомизированного исследования AMIRI—CABG (ClinicalTrials.gov Identifier: NCT03050489) в программном пакете SAS Enterprise Guide 6.1, позволяющего выполнить корреляционный анализ.

Материалы и методы

В проспективное нерандомизированное исследование AMIRI—CABG (ClinicalTrials.gov Identifier: NCT03050489) в Научно-исследовательском центре сердечно-сосудистой хирургии ПСПбГМУ им. И.П. Павлова с 2016 по 2019 г. включено 336 пациентов с ишемической болезнью сердца, имеющих показания к операции коронарной реваскуляризации. Создана база данных с результатами клинических, лабораторных и инструментальных исследований. Статистическая обработка базы данных

проведена с использованием лицензионного программного обеспечения SAS Enterprise Guide 6.1.

Результаты и их обсуждение

Как и в ранее рассмотренных примерах, прежде чем выполнять корреляционный анализ, необходимо определить тип распределения изучаемых переменных [1]. Если распределение отличается от нормального, стоит попробовать привести его к нормальному распределению. Как это сделать, было рассмотрено ранее [1]. В нашем случае к нормальному виду удалось привести распределение переменной TnIEndOp (уровень тропонина I к концу операции) с помощью логарифмирования. Переменные TnI1 (уровень тропонина I на первые сутки после операции) и AoClamp (время пережатия аорты) привести к нормальному распределению не удалось. В случае нормально распределенных переменных и предположения о линейной связи между переменными корреляцию рассчитывали с помощью коэффициента Пирсона. Если распределение переменных отличается от нормального, то следует использовать коэффициент корреляции Спирмена. Следует отметить, что коэффициент корреляции Пирсона позволяет выявить линейную связь. Если связь носит нелинейный характер, то данный метод покажет отсутствие корреляции [2].

Далее представлен код, с помощью которого можно определить характер распределения, а также привести распределение к нормальному, если оно отличается от нормального.

```
ods graphics on;
/* Выполняем проверку на нормальность
распределения по критериям Колмогорова -
Смирнова или Шапиро - Уилка, если p>0,05
то распределение нормальное */
Proc UNIVARIATE DATA=WORK.'20_06_2019
work'n normaltest plots;
  where CPBType=2;
  VAR TnIEndOp TnI1 AoClamp;
run;
/* приведем к нормальному распределению
с помощью логарифмирования */
DATA NEWDATASET; /* Создаем новую
таблицу данных */
  SET WORK.'20_06_2019 work'n;
/* переносим в новую таблицу все наши
данные */
  LGTnI1=LOG10(TnI1); /* логарифмируем
с целью "нормализации" */
  LGTnIEndOp=LOG10(TnIEndOp);
  LGAoClamp=LOG(AoClamp);
RUN;
/* TnIEndOp - уровень тропонина
к концу операции
TnI1 - уровень тропонина на 1-е сут-
ки после операции
AoClamp - время пережатия аорты; */
```

После определения характера распределения переменных построим скатерограмму (рис. 1). Оценим, существует ли корреляция между повышением уровня тропонина I к концу операции и на первые сутки после операции.

```
/* Оценим характер связи между уровнем тропонина I к концу операции и уровнем тропонина I на 1-е сутки после операции */
proc sgplot data=NEWDATASET;
  WHERE CPBTYPE=2;
  title "Скатерограмма»;
  scatter x=LGTnIEndOp y=LGTnI1;
  ellipse x=LGTnIEndOp y=LGTnI1;
  label LGTnIEndOp = 'уровень тропонина I к концу операции';
  label LGTnI1 = 'уровень тропонина I на 1-е сутки после операции';
run;
```

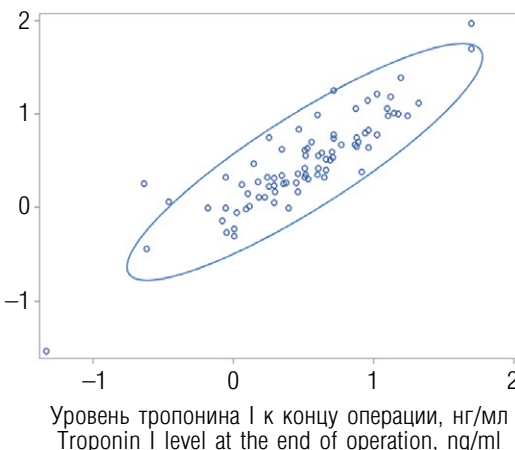
На рис. 1 видно, что точки стремятся распределиться вдоль наклонной прямой, следовательно, зависимость двух переменных носит линейный характер и можно применять линейный корреляционный анализ.

```
TITLE 'Корреляция исходных значений тропонина I к концу операции и на 1-е сутки после операции';
proc corr DATA=WORK.'20_06_2019 work'n
  pearson spearman kendall hoeffding fisher;
  WHERE CPBTYPE=2;
  var TnI1;
  with TnIEndOp;
run;
```

Примечания.

pearson — указываем, чтобы рассчитать коэффициент корреляции Пирсона (две исследуемые переменные распределены нормально);

Уровень тропонина I на 1-е сутки после операции, нг/мл
Troponin I level on the 1st postoperative day, ng/ml



- Уровень тропонина I к концу операции, нг/мл
Troponin I level at the end of operation, ng/ml
- 95 % доверительный эллипс
95% prediction ellipse

Рис. 1. Скатерограмма зависимости уровня тропонина I на первые сутки после операции от уровня тропонина I к концу операции

Fig. 1. Scatter plot for correlation analysis of troponin I level at the end of operation and on the 1st postoperative day

spearman — указываем, чтобы рассчитать коэффициент корреляции Спирмена (одна или две исследуемые переменные имеют распределение, отличающееся от нормального);

kendall — указываем, чтобы рассчитать коэффициент корреляции Кендалла;

fisher — указываем, чтобы рассчитать доверительный интервал корреляции.

Нажимаем кнопку «**Выполнить**» (рис. 2).

Получаем результаты, обращаем внимание на следующую таблицу (рис. 3).

```
ods graphics on;
/*
  Выполняем проверку на нормальность распределения по
  Критериям Колмогорова-Смирнова или Шапиро-Уилка, если p>0,05 то распределение нормальное
*/
proc UNIVARIATE DATA=WORK.'20_06_2019 work'n normaltest plots;
  where CPBType=2;
  VAR TnIEndOp TnI1 AoClamp;
run;
/*
  приведем к нормальному распределению с помощью логарифмирования
*/
DATA NEWDATASET; /* Создаем новую таблицу данных */
SET WORK.'20_06_2019 work'n; /* переносим в новую таблицу все наши данные */
LGTnI1=LOG10(TnI1); /* логарифмируем */
LGTnIEndOp=LOG10(TnIEndOp);
LGAoClamp=LOG(AoClamp);
RUN;
/*
  TnIEndOp - уровень тропонина к концу операции
  TnI1 - уровень тропонина на 1 сутки после операции
  AoClamp - время пережатия аорты;
*/
Оценим характер связи между уровнем тропонина I к концу операции и уровнем тропонина I на 1 сутки
proc sgplot data=WORK.'20_06_2019 work'n ;
  WHERE CPBTYPE=2;
  title "Скатерограмма";
  scatter x=TnIEndOp y=TnI1;
  ellipse x=TnIEndOp y=TnI1;
  label TnIEndOp = 'уровень тропонина I к концу операции';
  label TnI1 = 'уровень тропонина I на 1-е сутки после операции';
run;
/*
  Рассчитаем корреляцию уровня тропонина I к концу операции и на 1 сутки после операции
*/
TITLE 'Корреляция исходных значений тропонина I к концу операции и на 1 сутки после операции';
proc corr DATA=WORK.'20_06_2019 work'n pearson spearman kendall hoeffding fisher;
  WHERE CPBTYPE=2;
  var TnI1;
  with TnIEndOp;
```

Рис. 2. Расчет корреляции

Fig. 2. Correlation analysis

Pearson Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
TnI1	TnIEndOp	81	0.90346	1.49072	0.00565	0.90241	0.851930	0.936278	<.0001

Spearman Correlation Statistics (Fisher's z Transformation)									
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits		p Value for H0:Rho=0
TnI1	TnIEndOp	81	0.87681	1.36179	0.00548	0.87553	0.812515	0.918325	<.0001

Рис. 3. Результаты корреляционного анализа

Fig. 3. Correlation analysis results

Поскольку одна из переменных имеет распределение, отличающееся от нормального, то берем результаты из раздела Spearman Correlation Statistic (см. рис. 3). Значение корреляции — 0,88 (сильная корреляция), 95 % доверительный интервал — 0,81–0,92, уровень значимости $p < 0,0001$. Коэффициент детерминации $0,88^2 = 0,77$. Это означает, что одна переменная способна объяснить 77 % вариативности другой, что свидетельствует о сильной связи двух переменных.

Таким образом, уровень тропонина I на первые сутки после операции сильно коррелирует с уровнем тропонина I к концу операции. Безусловно, исходя из патофизиологии ишемически-реперфузионного повреждения миокарда, полученный вывод предполагался, но с целью демонстрации возможностей SAS и наглядности был выбран анализ двух переменных, между которыми, очевидно, существует связь.

Теперь выполним корреляционный анализ с целью выявления связи между длительностью пережатия аорты (AoClamp) и уровнем тропо-

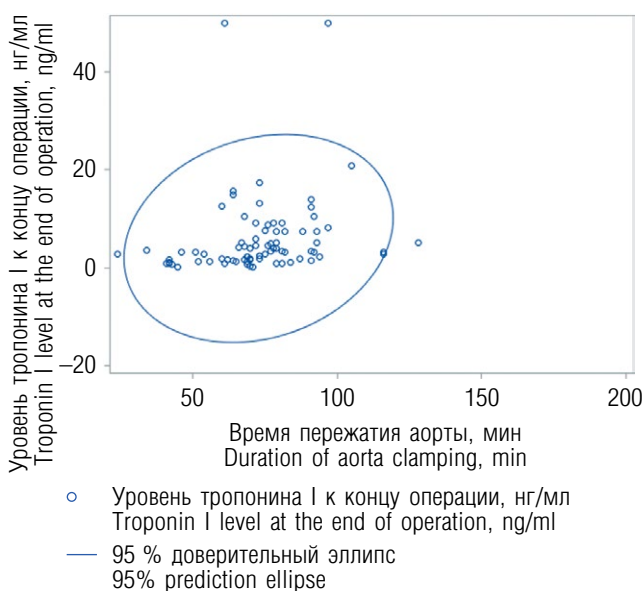


Рис. 4. Скатерограмма между временем пережатия аорты и уровнем тропонина I к концу операции

Fig. 4. Scatter plot aorta clamping time – troponin I level

нина I к концу операции (TnIEndOp). Строим скатерограмму:

```
TITLE 'Зависимость уровня тропонина I
от времени пережатия аорты';
proc sgplot data=NEWDATASET;
  WHERE CPBType=2;
  title "Скатерограмма";
  scatter x=AoClamp y=LGTnIEndOp;
  ellipse x=AoClamp y=LGTnIEndOp;
  label AoClamp = 'время пережатия
аорты';
  label TnIEndOp = 'уровень тропо-
нина I к концу операции';
run;
```

После нажатия кнопки «Выполнить» получим скатерограмму (рис. 4) зависимости уровня тропонина I к концу операции от времени пережатия аорты.

Обращает на себя внимание широкий разброс точек от границ эллипса: достаточно сложно провести прямую, вдоль которой могло бы расположиться большинство точек, эллипс стремится к контуру окружности (см. рис. 4). Уже по скатерограмме понятно, что корреляция между временем пережатия аорты и степенью повышения тропонина I концу операции отсутствует.

Рассчитаем коэффициент корреляции Спирмена.

```
TITLE 'Корреляция уровня тропонина I
к концу операции и времени пережатия
аорты';
proc corr DATA=NEWDATASET pearson
spearman kendall hoeffding fisher;
  where CPBTYPE=2;
  var AoClamp;
  with LGTnIEndOp;
  label LGTnIEndOp = 'уровень тропо-
нина I к концу операции';
run;
```

Полученный результат представлен на рис. 5.

Корреляция между временем пережатия аорты и уровнем тропонина I к концу операции — 0,4 (слабая), 95 % доверительный интервал — 0,19–0,57 — относительно широкий, уровень значимости $p = 0,0002$ (корреляция статистически значима). Коэффициент детерминации $0,4^2 = 0,16$. Исходя из расчетов, одна перемен-

Pearson Correlation Statistics (Fisher's z Transformation)								
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits	p Value for H0:Rho=0
AoClamp	LGTnlEndOp	80	0.45889	0.49591	0.00290	0.45660	0.263293 0.614650	<.0001

Spearman Correlation Statistics (Fisher's z Transformation)								
Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate	95% Confidence Limits	p Value for H0:Rho=0
AoClamp	LGTnlEndOp	80	0.39768	0.42089	0.00252	0.39556	0.192582 0.566080	0.0002

Рис. 5. Корреляция между временем пережатия аорты и уровнем тропонина I к концу операции

Fig. 5. Correlation analysis aorta clamping time – troponin I level

ная объясняет только 16 % вариабельности другой переменной, что крайне мало. Таким образом, можно заключить, что время пережатия аорты не влияет на повышение уровня тропонина I к концу операции. В случае коэффициента корреляции $r < 0,4$ говорят о слабой связи, при $0,4 < r < 0,8$ — о связи средней силы, при $r > 0,8$ — о сильной связи.

Как уже упоминалось, отсутствие корреляции может говорить об отсутствии линейной связи двух переменных, но возможна нелинейная связь. Может быть, она присутствует в нашем случае? На этот вопрос помогает ответить скатерограмма, представленная на рис. 4. Разброс точек стремится к равномерности, что говорит об отсутствии не только линейной связи, но и нелинейной. В случае нелинейной зависимости корреляционный анализ покажет отсутствие корреляции, поэтому при поиске зависимости между двумя переменными следует строить скатерограммы. Пример нелинейной зависимости между двумя переменными рассмотрен на рис. 6.

Следует отметить, что корреляционный анализ широко упоминается в российской литературе, но для понимания границ его применимости и интерпретации результатов необходимы знания биостатистики [2].

Несмотря на отсутствие корреляции, нельзя исключить наличия связи между исследуемыми переменными. Возможно, стоит одну из исследуемых переменных попробовать привести к бинарному виду и применить логистическую регрессию.

Разработанный алгоритм корреляционного анализа представлен таким образом, чтобы врач, никогда не работавший с программным продуктом SAS Enterprise Guide 6.1, мог обработать свою базу данных. Одно из преимуществ SAS — возможность достаточно просто рассчитать доверительный интервал для коэффициента корреляции, следует только указать в процедуре корреляционного анализа **fisher**, в то время как в других программных продуктах такой возможности нет и доверительные интервалы для коэффициентов корреляции необходимо рассчитывать вручную [2]. В данной

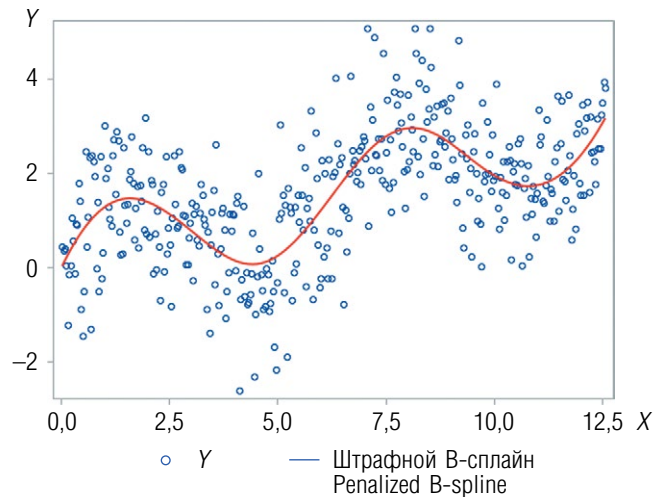


Рис. 6. Нелинейная связь между двумя переменными

Fig. 6. Nonlinear relationship

статье рассмотрены вопросы практического применения корреляционного анализа в программном пакете SAS, теоретические аспекты затронуты лишь в малой степени, так как прекрасно представлены в российской литературе [2–5].

Таким образом, SAS Enterprise Guide 6.1 предоставляет полный набор современных методов обработки данных, необходимых врачу-исследователю.

Выводы

1. SAS Enterprise Guide 6.1 позволяет быстро и удобно выполнять корреляционный анализ, что делает этот программный пакет интересным для врачей.
2. Разработанный алгоритм корреляционного анализа может применяться исследователями для обработки различных баз данных научных и клинических исследований.

Дополнительная информация

Финансирование. Работа выполнена в рамках государственного задания по теме «Оценка регенеративного потенциала пациента при операциях на сердце».

Соблюдение этических норм. Выполнение исследования одобрено протоколом локального этического комитета ФГБОУ ВО ПСПбГМУ им. И.П. Павлова.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Литература

1. Гржибовский А.М. Корреляционный анализ // Экология человека. – 2008. – № 9. – С. 50–60. [Grjibovski AM. Correlation analysis. *Ecology, human*. 2008;(9):50-60. (In Russ.)]
2. Гржибовский А.М., Иванов С.В., Горбатова М.А. Корреляционный анализ данных с использованием программного обеспечения STATISTICA и SPSS // Наука и здравоохранение. – 2017. – № 1. – С. 7–36. [Grjibovski AM, Ivanov SV, Gorbatova MA. Correlation analysis of data using statistica and spss software. *Nauka i zdravo-okhranenie*. 2017;(1):7-36. (In Russ.)]
3. Гржибовский А.М., Иванов С.В., Горбатова М.А. Экологические (корреляционные) исследования в здравоохранении // Наука и здравоохранение. – 2015. – № 5. – С. 5–18. [Grjibovski AM, Ivanov SV, Gorbatova MA. Ecological (correlation) studies in health sciences. *Nauka i zdravookhranenie*. 2015;(5):5-18. (In Russ.)]
4. Унгурияну Т.Н., Гржибовский А.М. Корреляционный анализ с использованием пакета статистических программ STATA // Экология человека. – 2014. – Т. 9. – С. 60–64. [Unguryanu TN, Grjibovski AM. Correlation analysis using STATA. *Ecology, human*. 2014;9:60-64. (In Russ.)]

Сведения об авторах / Information about the authors

Николай Сергеевич Буненков — аспирант кафедры факультетской хирургии, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. <https://orcid.org/0000-0003-4331-028X>. E-mail: bunenkov2006@gmail.com.

Nikolay S. Bunenkov — aspirant, Department of Faculty Surgery, First Pavlov State Medical University, Saint Petersburg, Russia. <https://orcid.org/0000-0003-4331-028X>. E-mail: bunenkov2006@gmail.com.

Гулнара Физулиевна Буненкова — клинический ординатор кафедры госпитальной терапии, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. E-mail: gulnara533@gmail.com.

Gulnara F. Bunenkova — resident, Department of Hospital Therapy, First Pavlov State Medical University, Saint Petersburg, Russia. E-mail: gulnara533@gmail.com.

Владимир Владимирович Комок — канд. мед. наук, кардиохирург отделения кардиохирургии № 2, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. <https://orcid.org/0000-0002-3834-7566>. E-mail: vladimir_komok@mail.ru.

Vladimir V. Komok — PhD, Cardiac Surgeon, Department of Cardiac Surgery No. 2, First Pavlov State Medical University, Saint Petersburg, Russia. <https://orcid.org/0000-0002-3834-7566>. E-mail: vladimir_komok@mail.ru.

Олег Александрович Гриненко — д-р мед. наук, проректор по лечебной работе, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. E-mail: klinika@spb-gmu.ru.

Oleg A. Grinenko — Doctor of Medical Science, Vice-rector, First Pavlov State Medical University, Saint Petersburg, Russia. E-mail: klinika@spb-gmu.ru.

Александр Сергеевич Немков — д-р мед. наук, профессор, кардиохирург, руководитель отделения кардиохирургии № 2, ФГБОУ ВО «Первый Санкт-Петербургский медицинский университет им. акад. И.П. Павлова» Минздрава России, Санкт-Петербург. <https://orcid.org/0000-0002-5152-0001>. E-mail: nemk_as@mail.ru.

Alexander S. Nemkov — Doctor of Medical Science, Professor, Cardiac Surgeon, Chief of Department of Cardiac Surgery No. 2, First Pavlov State Medical University, Saint Petersburg, Russia. <https://orcid.org/0000-0002-5152-0001>. E-mail: nemk_as@mail.ru.

✉ Контактное лицо / Corresponding author

Николай Сергеевич Буненков / Nikolay S. Bunenkov
E-mail: bunenkov2006@gmail.com